(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 February 2001 (08.02.2001)

PCT

(10) International Publication Number
WO 01/09155 A1

(51) International Patent Classification[7]: C07H 21/02, 21/04, C12N 9/00, 15/00, 5/00, 1/20

(21) International Application Number: PCT/US00/20331

(22) International Filing Date: 26 July 2000 (26.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/146,699 30 July 1999 (30.07.1999) US

(71) Applicant *(for all designated States except US)*: **MERCK & CO., INC.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MACNEIL, Douglas, J.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **OCCI, James** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US). **GEWAIN, Keith, M.** [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(74) Common Representative: **MERCK & CO., INC.**; 126 East Lincoln Avenue, Rahway, NJ 07065-0907 (US).

(81) Designated States *(national)*: CA, JP, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/09155 A1

(54) Title: CLONING OF THE *STREPTOMYCES AVERMITILIS* GENES FOR GLYCOSYLATION OF AVERMECTIN AGLYCONES

(57) Abstract: A cluster of genes involved in the synthesis and/or addition of oleandrose to avermectin aglycones has been cloned. A 11-kb *Pst*I clone complemented 28 avermectin glycosylation mutants in seven complementation classes. Sequencing of an 10-kb region identified 9 ORFs and an additional partial ORF. Eight of the ORFs were correlated to the seven glycosylation complementation classes. Sequence comparison to Genbank databases identified 6 genes: dTDP-glucose synthase; dTDP-glucose 4,6 dehydrase; dTDP-4-keto-hexose reductase; dTDP-hexose 3.5 epimerase; dTDP-hexose 3' O-methylase; and an avermectin aglycone-dTDP-oleandrose glycosyltransferase. The ninth ORF was essential for biosynthesis of the avermectin aglycones. The partial ORF encoded part of an avermectin polyketide synthase module 7.

TITLE OF THE INVENTION

CLONING OF THE *STREPTOMYCES AVERMITILIS* GENES FOR GLYCOSYLATION OF AVERMECTIN AGLYCONES

CROSS-REFERENCE TO RELATED APPLICATIONS

Not applicable.

STATEMENT REGARDING FEDERALLY-SPONSORED R&D

Not applicable.

REFERENCE TO MICROFICHE APPENDIX

Not applicable.

FIELD OF THE INVENTION

The invention is in the field of the genetics of biocatalysis and biosynthesis of secondary metabolites.

BACKGROUND OF THE INVENTION

Streptomyces are gram positive bacteria which undergo temporal differentiation from substrate mycelia to aerial mycelia and, later to, spores. *Streptomyces* produce a wide variety of secondary metabolites, including most of the known antibiotics. In order to better understand the biology of secondary metabolism, many genetic techniques have been developed for *Streptomyces* (reviewed by Hopwood, 1967; Chater and Hopwood, 1984). In addition, in order to isolate and study the function of *Streptomyces* genes involved in antibiotic production, recombinant DNA procedures have been developed (Hopwood et al., 1985).

The commercially important Streptomycete, *S. avermitilis* , produces a series of eight related oleandrose containing, polyketide macrolides, termed the avermectins (Burg *et al.*, 1979). Avermectins are potent anthelmintic compounds which are active against many endoparasites of animals and humans, including *Onchocerca volvulus* the agent of "river blindness". The avermectins are also active against arthropod ectoparasites (Fisher *et al.*, 1984) and are effective in controlling numerous agricultural pests (Putter *et al.*, 1981). The semi-synthetic avermectin, ivermectin, is a major compound in use world wide for control of animal parasites.

Therefore, it is commercially important to know how many genes are involved in the biosynthesis of the avermectins, how the genes are regulated, and what the genes' functions are. Efficient procedures for transformation of *S. avermitilis* have been developed (Klapko & MacNeil, 1987) and a variety of plasmid vectors have been identified which replicate in *S. avermitilis* (Klapko & MacNeil, 1987; MacNeil, 1988; MacNeil & Gibbons, 1986).

Mutants of *S. avermitilis* that have altered pathways of avermectin biosynthesis have been described. These includes a mutant which fails to close the furan ring of avermectin (Gogelman *et al.*, 1983), a mutant which produces avermectin aglycones (Schulman *et al.*, 1985), and mutants which are deficient in O-methylation of avermectin (Ruby *et al.*, 1986; Schulman *et al.*, 1987). Ikeda *et al.* (1987) reported the isolation of two classes of *S. avermitilis* mutants. These include nonproducers (NPA mutants), which produce no detectable avermectins; aglycone producers (AGL mutants), which are blocked in the glycosylation avermectin aglycones; OMT mutants which lack the ability to methylate the O at C-5, and GMT mutants which lack the ability to methylate the O at C-3' and C-3" of the oleandrose moiety. Ikeda *et al.* used a natural fertility system to show linkage between the mutations in these classes, indicating that at least some of the genes for avermectin biosynthesis are clustered.

The genes for avermectin 5-keto reductase and avermectin 5 O-methyl transferase have been cloned (Ikeda *et al.*, 1995; Ikeda *et al.*, 1998). A series of overlapping cosmid clones representing 150 kb of genomic DNA were isolated by complementation of C-5 O-methyl transferase mutant and glycosylation deficient mutants. Deletion mapping over a 150 kb region located the avermectin gene cluster to a 100 kb segment (MacNeil et al., (1993). Complementation analysis, using various restriction fragments from one end of the avermectin gene cluster, has identified 3 complementation classes involved in the synthesis and/or attachment of oleandrose to the avermectin aglycone (MacNeil *et al.*, (1992)).

SUMMARY OF THE INVENTION

The present invention extends the genetic analysis of the avermectin genes involved in glycosylation. Through sequencing and analysis of a 10 kb segment of the genome of *Streptomyces avermitilis* the invention provides polynucleotides of eight ORFs that correlate to seven glycosylation deficiency complementation classes. The invention further provides eight polypeptides encoded by the ORFs.

Aspects of this invention are isolated nucleic acid fragments of the 11 kb fragment of the *S. avermitilis* genome disclosed herein. The fragments preferable encode at least one of the proteins encoded on the genomic fragment. Any such polynucleotide includes but is not necessarily limited to nucleotide substitutions, deletions, additions, amino-terminal truncations and carboxy-terminal truncations such that these mutations encode an ORF that can be expressed as a protein or protein fragment of enzymatic, biochemical, biosynthetic or diagnostic use.

In particular embodiments, the isolated nucleic acid molecule of the present invention can be a deoxyribonucleic acid molecule (DNA), such as genomic DNA and complementary DNA (cDNA), which can be single (coding or noncoding strand) or double stranded, as well as synthetic DNA, such as a synthesized, single stranded polynucleotide. The isolated nucleic acid molecule of the present invention can also be a ribonucleic acid molecule (RNA). In particular embodiments, the nucleic acid can include the entire sequence of the gene cluster, the sequence of any one of the ORFs, a sequence encoding an ORF and an associated promoter, or smaller sequences useful for expressing peptides, polypeptides or full length proteins encoded in the fragment of the *S. avermitilis* genome disclosed herein. In particular embodiments the nucleic acid can have natural, non-natural or modified nucleotides or internucleotide linkages or mixtures of these.

Aspects of the present invention include nucleotide probes and primers derived from the nucleotide disclosed herein. In particular embodiments of the invention, probes and primers are used to identify or isolate polynucleotides encoding the avermectin pathway proteins disclosed herein or mutant or polymorphic forms of the proteins. Probe and primers can be highly specific for the nucleotide sequences disclosed herein.

An aspect of this invention is a substantially purified form of a protein described herein. In preferred embodiments the proteins have the amino acid sequence disclosed herein and set forth in SEQ ID NOs.

Aspects of the present invention include fragments, polymorphs and/or mutants of the polypeptides disclosed herein, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these mutations provide for active proteins or active protein fragments or protein fragments of diagnostic use.

Aspects of the present invention include recombinant vectors and recombinant hosts which contain the nucleic acid molecules disclosed throughout this

specification. In particular embodiments, the vectors and hosts can be prokaryotic or eukaryotic. In particular embodiments the hosts express peptides, polypeptides, proteins or fusion proteins of the avermectin pathway polypeptides disclosed herein. In further embodiments the host cells are used as a source of expression products.

Aspects of the invention are polyclonal and monoclonal antibodies raised in response to either the entirety of a polypeptide disclosed herein, or only a fragment, or a single epitope thereof.

Aspects of this invention include the use of the nucleic acids or proteins disclosed herein, and their active polypeptide fragments, together, individually, or in combination with other enzymatically active polypeptides to perform combinatorial biocatalysis *in vitro* and *in vivo* in an appropriate host cell. In preferred embodiments, the nucleic acids or polypeptides disclosed herein are used to perform biotransformations of macrolide compounds, including the glycosylation of avermectin or other macrolide aglycones. In particular embodiments, the nucleic acid and proteins can be used *in vivo* in a bacterial host, *in vitro* in combination with an actinomycete fermentation, or *in vitro* in combination with enzymatically active polypeptides that are not from the avermectin biosynthetic pathway to effect the synthesis of a pharmaceutically active compound, including but not limited to an antibiotic compound.

Each document mentioned in this specification is hereby incorporated herein by reference in its entirety.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1. A map showing the location of the 8 avermectin genes on the 11 kb PstI fragment and indicating the subclones of the region used in the complementation analysis.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides nucleotide sequences of eight genes of the *Streptomyces avermitilis* avermectin biosynthesis pathway. The genes are a cluster of genes involved in the synthesis and addition of oleandrose to avermectin aglycone. The invention also provides the polypeptides encoded by these genes. The genes and polypeptides can be used to glycosylate avermectin aglycones, other macrolides or other hydroxy compounds. The genes and polypeptides can be used in combination with other biosynthetic genes to produce known or novel compounds.

Polynucleotides

A preferred aspect of the present invention is a nucleic acid that encodes at least one polypeptide encoded by the sequence disclosed below. A preferred embodiment is a nucleic acid that encodes at least one polypeptide encoded by the sequence disclosed below and has the same sequence from that segment of the sequence disclosed below follows:

1 GGATCCATCG CCAACGCCTC ACGCGGACTG ATCCCGAAAA ACCCCGCATC

51 GAACTCCGCC GCACCCTCCA GGAAACCGCC CCGGCGCGTA TACGACGAAC

101 CCGCCCGCCC CGGCTCCGGA TCATAGAAAG CCTCCACGTC CCAACCCCGG

151 TCGACCGGAA ACTCCCCCAC CGCATCCCGA CCCGACGCAA TCAACTCCCA

201 GAAATCCTCC GCCGACTCCA CACCCCCCGG AAAACGGCAC GCCATCCCCA

251 CAATTGCAAT CGGCTCCTGC TCGCCCGATT CAATCTGCTG AAGTCGACGC

301 CGCACATTGA GGAGATCGGC AGTAACGCGC TTGAGATAGT CGCGGAGCTT

351 TTCCTCGTTA GCCATGGACC GGTCTCCTCG ACAAGAGAAA TCGGAAATTA

401 AAAAACACGC ATGGGACTCT CACAGGCTAG AGCGACGAGA GCAGCACAAA

451 TACCCCTAGA TACCCCAGAC CCCTGATGCT CGATGAATGC CGCTATAGCT

501 AGGGGGTATG GCGCCAGACA TGAATTCACA GCGTTTCGGC GGCCGGCTGG

551 CGCTTGTCAC AGGTGCAGGC GGTGGCATCG GGCGGGCGAC CTGCGCTCTC

601 GGATCGGCCG GGGCGCGAGT GGTGTGCGTG GACCGGGACG GCCGCGGCGC

651 CGGGGTGACG CCGACCTGGC CGGAGCGGGG CGCGCGGGCG GCCTGGCCCG

701 AGGTGGCCGA CGTGTCCGAC GGAGCGGCGA TGGAGCGGTT GCCCGAGCGC

751 GTCGCCGAGA CGTACGGGGT CGTGGACCTG CTGGTGAACA ACGCCGGCAT

801 CGGCATGGCG GGGCGTTTTC TCGACACGTC CGTCGAGGAC TGGCAGCGCA

851 CCCTGGGCGT CAACCTCTGG GGTGTCATTC ATGGTTGCCG CCTCATCGGC

901 CGGCAGATGG CGGAGCGCGG GCAGGGCGGG CACATCGTGA CGGTGGCGTC

951 GGCGGCGGCG TTCCAGCCGA CGCGGGCGGT CCCCGCGTAT GCCACCAGCA

1001 AGGCGGCGGT GCTGATGCTG AGCGAGTGCC TGCGCGCGGA GTTCGCGGAG

1051 TTCGGGGTCG GAGTGAGCGT GGTGTGCCCG GGCTTCGTCC GTACGTCGTT

1101 CGCGTCGGCG ATGCATTTCG CCGGTGTGCC CCGGCTGGAG CAGGAGCGGC

1151 TGCGGGCGCT GTTCGCCGGT CGCGGATGCA GCGCGGAGAA GGTGGCCGCG

1201 GCGGTACTGC GGTCGGTGGC GCGCGACTCG GCCGTGGTGA CCGTGACGGC

1251 GGAAGCGCGG CTGTCACGGC TGATGAGCCG CTTCACGCCA CGCCTGCGCG

1301 CCGCGGTGGC GCGGATGGAT CCCCCTTCGT AGGGCTGGCG GGGATCCCCT

1351 CCTTGCCTTC GAACATCTTC CGACGATGGG CAGTGAGAGA TGTCAGATCA

1401 TTTTCTCTTC ATGAGTGCGC CGTTCTGGGG GCATGTGTTC CCCAGTCTCG

1451 CCGTGGCGGA GGAGCTCGTG CACCGGGGCC ACCACGTCAC CTTTGTGACG

1501 GGCGCGGAAA TGGCCGATGC GGTGCGTTCC GTGGGCGCTG ATTTCCTGCG

1551 GTACGAGTCC GCCTTCGAGG GTGTCGACAT GTACCGGCTG ATGACCGAGG

1601 CCGAGCCGAA CGCCATCCCC ATGACGCTGT ACGACGAGGG CATGTCCATG

1651 TTGCGTTCGG TGGAGGAGCA CGTCGGCAAG GACGTTCCGG ACCTGGTGGC

1701 CTACGACATC GCCACCTCCC TCAACGTGGG TCGTGTCCTC GCCGCCTCCT

1751 GGAGCAGGCC GGCCATGACG GTCATTCCCC TGTTCGCGTC CAACGGGCGC

1801 TTCTCCACGA TGCAGTCGGT ATTGGATCCG GATTCCGCTC AGGTCAGTGC

1851 GCCGCCGCCG CGCTTCTCGG AGCAGATGGA GTTGTTCGGC CTCGGGGCGC

1901 TGGTGCCGCG CCTCGCGGAG CTGCTCGTTT CCCGGGGTAT CACGGAACCG

1951 GTCGACGATT TCCTTTCCGG ACCGGAGGAC TTCAACCTGG TGTGTCTGCC

2001 GCGCGCCTTC CAGTACGCGG GCGACACCTT CGACGAGCGG TTCGCCTTCG

2051 TCGGACCATG TCTGGGTAAG CGCAGGGGTC TGGGCGAGTG GACACCACCG

2101 GGCAGCGGGC ATCCAGTGGT GCTCATCTCC CTCGGGACCG TGTTCAACCG

2151 GCAGCTGTCC TTCTTCCGCA CGTTCGTCCG GGCGTTCACC GACGTCCCCG

2201 TGCACGTCGT GATCTCGCTC GGCAAGGGGG TCGACCCCGA TGTGCTGCGG

2251 CCGCTGCCGC CGAATGTCGA GGTGCACCGG TGGGTGCCGC ACCATGCGGT

2301 GCTGGAGCAT GCCAGGGCTC TGGTCACGCA CGGCGGTACC GGCAGTGTGA

2351 TGGAGGCACT GCACGCAGGG TGCCCGGTGC TCGTCATGCC CTTGTCGCGG

2401 GACGCGCAGG TGACCGGCCG GCGGATCGCC GAGCTGGGGC TGGGTCGTAT

2451 GGTGCAGCCG GAGGAGGTCA CGGCGACGAC GCTGCGCCGG CACGTGCTGG

2501 ACATCATCTC CGATGACGCG ATCACCCGAC AGGTCAGGCA GATGCAGCGG

2551 GCCACGGTCG AGGCGGGCGG CGCCCTGCGG GCAGCGGACG AGACCGAGCG

2601 GTTTCTGCGC CGGACGCGCC GTCACTGACC GGCAGCTCGG GCCGGGCGGT

2651 GAGTGGCTCC CACAGGGTTC GGTTCTCCAC GTACCACTGA ACGGTCTGTG

2701 CCAGCCCCTC CTCGAAGGGC ACGCGGGGCG CGTAACCGAG CTCGGCGGAG

2751 ATCTTGCTGA TGTCCAGCGA GTAGCGCCGG TCGTGCCCCT TGCGGTCGGT

2801 CACGGGTTCG ACCATCGACC AGTCCACGCC GAGCAGGTCC AGGAGCCGGG

2851 CGGTGAGCTC ACGGTTGGAC AGCTCCGTCC CGCCTCCGAT GTGGTAGATC

2901 TCGCCGGGCC TGTCGCGTTC GGCGACCAGG GCGATGCCAC GGCAGTGGTC

2951 GTCCACGTGC AGCCAGTCGC GGACGTTTTC GCCGTCGCCG TACAAGGGCA

3001 CCTTCGTGCC GTTCAGCAGA TGGGTGACGA ACCGCGGGAT GAGTTTCTCC

3051 GGGAACTGGT GGGGGCCGTA GTTGTTCGAG CATCGGGTGA TGATCACTGG

3101 TAGGCCGTGC GTGCGGTGGA AGGACCGGGC GAGCAGGTCG GAGGACGCCT

3151 TGGACGCGGA GTAGGGCGAG TTCGGCTCCA GCGGGGCGTC CTCGGTCCAC

3201 GAGCCGGAGT CGATGGAGCC GTAGACCTCG TCCGTCGAGA TGTACACGAA

3251 GCGGTCCACG GCGGCGTCGG TGGCGGCGCG GAGCAGGGTG TGAGTGCCGA

3301 GGACATTGGT GCGTACGAAC TCGGCGGCGT CGGCCACGGA CCGGTCCACG

3351 TGTGACTCCG CCGCGAAGTG GACCACCATG TCGGAGCCGT CCATCAGGTC

3401 CGCGACCAAG GGCCCGTCGC AGATGTCGCC GTGCACGAAG ATCAGGGATG

3451 GGCTTCCCAG GACCGGTGCG AGGTTCTCCA GGCGACCCGC GTAGGTCAGC

3501 TTGTCGAGCA CCACGACCTC GGCACCGGTG AACGCCGGAT ACGCGCCCGT

3551 CAGCAACCGC CGTACGAAAT GGGAACCGAT GAAACCGGCG CCGCCCGTCA

3601 CGAGTAGGCG CATCCCGGGC TCCTCACCGC GGCTTCCGCC GCAATACTCA

3651 TCAGATACTC GCCGTAGCCG GAGCCGGCCA GTTCGACCCC GCGCAGATAG

3701 CAGTCGTCCG CGTCGATCAG ACCCATCCGG AAGGCGATCT CCTCGAGACA

3751 GGCGATCCGT ACTCCCTGGC GCTTCTCCAG GACCTGCACA TACTGCCCGG

3801 CGTGCATCAG CGAGTCGTGC GTCCCCGCAT CGAGCCAGGT GAAGCCCCGG

3851 CCCAGGTCCA CCAGCCGGGC CCGCCCCTCG GCGAGGTAGG CCCTGTTGAC

3901 GTCGGTGATC TCCAGCTCGC CGCGGGCCGA CGAGCGGATG CCCCGGGCCA

3951 CCTCGATCAC GTCGTTGTCG TACAGGTACA GGCCTGTGAT CGCCAGGTTG

4001 GACCGGGGGG CGGTGGGTTT CTCCTCGACG GACAGCAGCT TTCCGGAGGC

4051 GTCGACCTCT CCGACTCCGT ACCGTTCGGG ATCCGTCACC GCGTATCCGA

4101 ACAACACACA GCCGTCGACA TCGCGGGTGT GGCTGCGCAG CAGGTGCGAA

4151 AAGCCCATGC CATGGAAGAT GTTGTCCCCA AGGACAAGGG ACACCTGATC

4201 CTGACCGATG AAATCGGCGC CGATGAGGAA TGCCTCGGCG ATTCCTCCCG

4251 GTCGCTGCTG CGCGGCGTAG TCGATGTTCA GCCCGAGGCG GCTTCCGTCT

4301 CCGAGCAGTC TCCGGAATTG TTCGAGATGA TCGGGTGAGG AAATCACCAG

4351 GATGTCTTTT ATGCCGCCGA GCATCAACAC GGAGAGCGGG TAGTAGATCA

4401 TGGGTTTGTC GTAGACAGGG AGCAGCTGCT TGGAAAGGGC ACGGGTCAAC

4451 GGGTAAAGCC GAGAGCCGGT TCCCCCCGCG AGCACGATTC CCTTCATGTC

4501 GGACTCCCCG CAGTCGACGT TATATATCTC GTGCCGTCTG CCCGACGGTA

4551 CCAAGTGGCG GAAAACGCAC CAGGAATTCG AGCGCCGCTA GGGGGAAGGG

4601 CTCAAGAAGA TAGGGGCCAC CAGATGGGGC GGTTTTCGGT GTGCCCGCCC

4651 CGGCCGACCG GAATACTGAA GAGCATGCTG ACGACTGGGA TGTGCGACCG

4701 ACCGCTGGTC GTCGTACTCG GAGCCTCCGG CTATATCGGG TCGGCCGTCG

4751 CGGCGGAACT CGCCCGGTGG CCGGTCCTGT TGCGGCTGGT GGCCCGGCGA

4801 CCGGGCGTCG TTCCGCCGGG CGGCGCCGCG GAGACCGAGA CGCGTACGGC

4851 CGACCTGACG GCGGCGAGCG AGGTCGCCCT CGCCGTGACG GACGCCGACG

4901 TGGTGATCCA CCTGGTCGCG CGCCTCACCC AGGGAGCGGC ATGGCGGGCG

4951 GCGGAGAGCG ATCCGGTGGC CGAGCGGGTG AACGTCGGGG TGATGCACGA

5001 CGTCGTCGCG GCCCTGCGGT CCGGGCGCCG CGCCGGGCCG CCCCCGGTGG

5051 TGGTGTTCGC CGGGTCGGTC TACCAGGTGG GCCGCCCGGG TCGGGTCGAC

5101 GGCAGTGAGC CGGACGAGCC CGTGACGGCC TATGCCCGTC AGAAACTCGA

5151 CGCCGAACGG ACGTTGAAGT CCGCCACGGT CGAGGGTGTC CTGCGGGGGA

5201 TCTCGCTGCG GCTGCCCACC GTCTACGGCG CGGGGCCGGG CCCGCAGGGC

5251 AACGGCGTCG TGCAGGCGAT GGTGCTCCGG GCGCTCGCCG ACGAGGCCCT

5301 CACCGTGTGG AACGGAAGCG TGGTGGAGCG TGACCTGGTG CATGTGGAGG

5351 ATGTCGCGCA GGCCTTCGTG AGCTGCCTGG CGCACGCGGA TGCGCTCGCC

5401 GGGCGGCACT GGCTGCTCGG CAGCGGTCGT CCTGTGACCG TCCCGCACCT

5451 CTTCGGTGCC ATCGCCGCCG GCGTGTCCGC CCGCACCGGG CGCCCCGCGG

5501 TGCCCGTGAC CGCGGTGGAC CCTCCGGCGA TGGCGACGGC GGCGGACTTC

5551 CACGGGACCG TCGTCGACTC CTCGGCGTTC CGCGCGGTCA CCGGGTGGCG

5601 GCCGCGGCTG TCGCTTCAGG AGGGCCTGGA CCACATGGTG GCGGCTTACG

5651 TGTAGCGCCG GGGTGGCGGC CGGGCCCGGG CGGTGACGGC CCGGATCCGG

5701 GTCGGCCGTC ACAGCTTCTC GTCGAGGGCG GGGCTCGCGC GGTACTCCGG

5751 CAACATGCCG CGTCGCAGGG CCTGCTGGAG AGTCGGCGCG CGCGCCGGTC

5801 CGCGCTCGGA GAGGATCGGT GCCCGCCCGA GGTGGTGGCC GAGGGGCAGG

5851 GCGAGGTCCG GATCCTCGGG CGAGAGGGCG TGTTCGTTCT GCGGAACGTA

5901 GCCGCTCGAC ATCAGGTACA CCATCGCCGT GTCGTCTTCC AGCGCCACGA

5951 ACGCGTGCCC GACCCCGATC GGCAGGTAGA CGGAACGGAA GCGCTCCTGG

6001 TCGAGGAGGA CCGAGTCCCA CTGCCCGAAA GTCGGTGAGC CGGTGCGCAG

6051 GTCGACGACG AAGTCCAGGG CCCGTCCCCG GGCGCAGTGG ACGTACTTGG

6101 CCTGGCCGGG TGGTGTCGCG GTGAAGTGCA CGCCGCGGAC GACGCCGCGG

6151 CGCGAGACGC TCTGGCAGGT CTGCGCGGTG GGAAACCGGT GCCCGACGGC

6201 CTCGCTGAGG ACCGGTTCCT GGTAGGGGGT GACGAAGAGC CCGCGCTCGT

6251 CGGGGAAGAC CGTCGGGGTG AATTCGACGG CGCCCTCGAC GACGAGCCTC

6301 CGGACCGTGA CACCGGCGGC GGTGGCCCGG GCGCCCGCGG GCGGGGCGGG

6351 CCGGTCGGCG GAGCTCCGGC GAGGCCGGCC AAGGGTCATC GCTGCACTCT

6401 CTCTGTCGTG CGGGTTGTCA TACGGGTAGT CGTACGGGCC GGTTCCGGAG

6451 TCACAGCTCG ACGGCGCGGG TGGTGAGCAG GGACAGCAGG GTGCGGGCCT

6501 GCACGTTCAC GTAACGGCCG TACCGCAGCA GCTGGGTCAG CTGGCCCGGG

6551 GTGCACCAGC GGTACCCCGG GGGCGGGTCG TTCGGCGCCT GGCTCTCGTC

6601 GGCCTCGACG AACAGGTAGC GCGCCTGTGC GTGCAGAAAG CGACCGCCCT

6651 CCTCCGAGTG GACCGCCGCG TAGCGGATGC GGTCGGGCGC GGCCTCCAGC

6701 ACCAGGTCGA GGAAGCGCGG CCTGGCCGGT CCCGTGAGGT GGGCGTAGTT

6751 GCGCGGGGTG TACTGGACCG TCGGGCCGAG TTCGATCGTG TCGAGGAAGC

6801 CGCCCTCGAC CCTGCCGTGG GCGAGCAGGT GCGGTACGCC GCCGATCCGC

6851 CGGGTCAGGA AGGCGGTGAT GCCGTGGCCG CACGGTTCGA TCAGGGGCTG

6901 GGTCCAGGCG GCGACCTCCC GGTTGGAGGC CTCGACACGG ACCGCGACCA

6951 CACGGAAGTA CCGGTCCGCG TGGTGGGCGA TGGACTCCGC GCCCGTGGTC

7001 CAGCCGGGGA TGCCGGCCAG GGGCACGCGG CGGGCGTGCA CGGAGTGCCG

7051 GGAGCGTTCG GCGGCGTACC AGGAGAGCAG TTCGGCGTCG CTGTGCAGGG

7101 CCGCGGGCTC GTCGAACGGG GTGGGAAGGC AGGCGAGGAC CGTGCGTGCG

7151 TCCATGTTCA CCAGGTTGTC CCGGTGCATC AGTTCGCCGA TCTGCCCCAG

7201 TGTCAGCCAG CGGAAGTCGT CGTCCAGTGG TACGTCCTCG TCGGTCTCCA

7251 CCACGATGTT GCGGTTGAAC TTCCGGTGGA ACCAGGCTCC GTGCTCGGAC

7301 TGGAGGACGT CGACCACCAC GGTGGCGCGC CGGGGCTGTG TGAAGTACTC

7351 GAGGTACTTC ACGGCGGCGC CCCCGTGGAC CTTGGTGTAG TTGCTGCGCG

7401 TGGCCTGCAC GGTGGGCGAC AGCTGGACCA GGTTGATGTT GCCGGGCTCC

7451 ATCTTGGCCT GCATCAGGAA GTGCAGGACC CCGTCGAACT TCTTGGCGAG

7501 GATGCCGAGG ATGCCGATCT CGGGCTGGTG GATGATGGGC TGCTGCCATT

7551 CCGGGAAGGG CTGTTCACCG CCTCGGACGT GCAGTCCCTC CACGGAGAAG

7601 AACCGGCCGC TGCGGTGGGC CAGATTGCCG GTTCCGGGGT GAAACGACCA

7651 GGCGTCCATC CCGTGGAAGG GGATGCGCTC GACCCGGAAC CGGTGGGCCC

7701 CGGACCGCCG CGTCCACCAG CCGGTGAACG CGTCGAGGGA CGTCCGGGCG

7751 CCGGTGTCGC CCACGGCGGC GGAGCGGGCG AGGCACGCGG GCAGGGCGGC

7801 GTCGTGCCGC GCGGTGAGCG GTGCTGGGCT CGGTGTGGTC GGCATCGGCT

7851 CGTACGCTCA TGCACCCCAC GTCATGTAGA TCACCGGTGG CTCGCGGCCG

7901 GGCAGTTGGC GCAGTGGGGC GTGGTCGAGG CCGAACGCCT CGCTCAGCGC

7951 CCTGGTCTCC CCCGGCCATT TGGGGTGGGT GAGTTCGTCG AAGGCGAGGA

8001 TGCTGCCCCT GGTCAGGTGC GGTGTGATGA CGTCCAGCAG TTCGCGCGTG

8051 GGGCGGTAGA GGTCCAGGTC GAAGTAGGCC AGCGCGATGA CGGTGTGCGG

8101 GTGTTCCGCC AGGTATTGGG GCACCGTTTC GCGTACGTCG CCCTGGACCA

8151 CGAAGGAACG CTGGGTGTGG CCGTAGGGTT CGTTCGCCTC GTGCGCCGCG

8201 AGCACCTGCC GCAGGTGCTC CACTTCGCCG TCCGGCACGG CGAACCGCCC

8251 AGGGACCGCG CTGGTGCTGA CCTCGTCCGC CTCGTCGATG TCGGGGAAGC

8301 CGGTGAACGT GTCGAAGCCG ATGACGCGGC GCAGCGAGTT GTACGGCTCA

8351 TAGATGCTGC GCAGCGCGGT CAGCGTGGCG AGGTGCCGTC CGTGCAGAAC

8401 GCCGAACTCC ATGATGACGC CGGGGACTTC CGGCAGCATG CGGTACAGCG

8451 CGTCCATGGA GAGCAGGTCG GCGAGCTGGT TGCGCCGCAT GTAGACGGAC

8501 AGGTTGTCGA TCAGGTACTT CGGCGGGATC GGGCTGTCGA CGAGGAGCTT

8551 GGTCAGCTGC TCGCGGGCAG CGCGTTCCTG CTCGGACTCG TGCGGCACGA

8601 TCCGGGGATC GGTGAACTCC CGCTCGGTCA TGGAGGCCTT TCCTTTCATG

8651 GGTCGGTACC GGGCGCGCCG GACGTGCCGG TCGTACCGGG CGTGCCGGCG

8701 GGCACGACGC TGTCGGGTCA GGACAGCCAG GCGTCGGGGG CGGATCCGCC

8751 GCGGCCGACC GGGGGGAACA GCTCCTCCAG GCGGGCCAGG ACGGGCTCGG

8801 GCAGCGGGGT GCGCAGGGCG TGCAGTGCCC CGTCCACGTG CTGTTCGGTG

8851 CGCGGCCCGA TGACCAGCCC GGTCACGCCG GGCCGCGACA GCACCCAGGC

8901 CATGCCGACA TGGGCGGGGT CGAGGCCGTG GTCCGCGCAC ACGTCCTCGT

8951 ACGCCGCGAT GGTGGTGCGG TGGTGCTCCA GGGCCTCGAC GGCCCGGCCC

9001 TGTGCCGACT TGACCGCGGT GTTCTCCCGC GTCTTGCGCA GGACACCGCC

9051 GAGCAGGCCG CCGTGCAGTG GCGACCAGAC CAGGACGCCG ACACCGTAGG

9101 CGGACGCGGC GGGGATGACT TCCAGCTCGG CGTGTCGGGT CACGAGGTTG

9151 TAGACGCACT GCTCGGAGGC GAGGCCCAGG GCGTTGCGCC GCCGGGCCGC

9201 CTCCTGGGCG GAAGCGATGT CCCAGCCCGC GAAGTTGGAG GAGCCGACGT

9251 AGCGCACCTT GCCCTGCGTG ATGAGCAGGT CCATCGCCTG CCACACCTCG

9301 TCCCAGCCGG CGCGGCGGTC GATGTGGTGC AGCTGGTACA GGTCGATCCA

9351 GTCGGTGCGC AGTCGGCGCA GCGAGGCGTC GCAGGCGGCC ACGATATTGC

9401 GTACGGACAG TCCGTGATCG TTGGGGCCGC TGCCCATCGG ATCGCCGACC

9451 TTGGTGGCCA GCACCACCTG CTCACGCCGG GCGGGGCGGT CCGCCAGCCA

9501 CCTGCCGATG ACCTCTTCGG TGTACCCCTT GTGGACGCGC CAGCCGTAGG

9551 TGTTGGCGGT GTCGAACAGG GTGATGCCCT GAGCCAGGGC GTGATCCATC

9601 AGTCGGCGCG CTTCGGGCTC CTCCACCCGT CCGCCGATGT TGACCGTTCC

9651 GAGCGCCAGT CGGCTGATCC TCAGCCGGGT CCTGCCCAGT TCGGTGTGGA

9701 GGGGAGCACT GCTGTTGCTG TCGGACTGGA CGGGTGCGGG CTCGGCCGTC

9751 GTAGGCATCA TCGATCAGTC GACACTCCCT CGTGCGTGAG CGGCGGGCGC

9801 TCGAGCAGGA CCCTGACCTG AGGCCCAGGA GGCTACCGGC GATCATGCGA

9851 TACAGGCAGC CGCTCGATGG TGGGACACGG GCTGCCGTCG CCGGGCATAG

9901 GGGCTGATGG GGGTTGTCCG GTGCGGGTCC GGCTGACAGC CTCGTGGACA

9951 CCAAGTTGAT CCAGTTGATC CACTCCGAAA GGCAGAGGCT GCAG
(SEQ ID NO:1)

The sequence SEQ ID NO:1 is characterized by the following open reading frames (ORFs) noted below. Each ORF encodes a protein in the avermectin biosynthetic pathway. Avermectin glycosylation genes AvrB, C, and D were identified by complementation analysis previously (MacNeil et al Gene (1992) 111:61-68 and map to ORF2, ORF3b and ORF3a respectively. Newly identified genes for avermectin glycosylation are designated AvrE, F, G, H, and I.

Mod7-PKS 1-365 - Beginning of mod7 PKS
ORF1 508-1332 + 274 aa Macrolide B-keto reductase
ORF2 AvrB 1390-2628 + 386 aa Glycosyl transferase
ORF3a AvrD 3598-4497 - 300 aa TDP-glucose synthase

ORF3b AvrC 3613-2534 - 360 aa TDP-glucose 4,6 dehydrase
ORF4 AvrE 4624-5655 + 343 aa Glycosyl reductase
ORF5 AvrF 5709-6389 - 226 aa Glycosyl 3,5epimerase
ORF6 AvrG 6451-7845 - 464 aa Oleandrose synthesis
ORF7 AvrH 7858-8631 - 257 aa Glycosyl methyltransferase
ORF8 AvrI 8718-9761 - 347 aa Oleandrose synthesis

Promoters:

1) Divergent PKS7-ORF1,2 between 365 and 508.
2) Divergent ORF3a,b-ORF4 between 4497 and 4624
3) ORF8,7,6,5 9994 to 9761

An isolated nucleic acid molecule of the present invention can include a deoxyribonucleic acid molecule (DNA), such as genomic DNA and complementary DNA (cDNA), which can be single (coding or noncoding strand) or double stranded, as well as synthetic DNA, such as a synthesized, single stranded polynucleotide. The isolated nucleic acid molecule of the present invention can also include a ribonucleic acid molecule (RNA).

As used herein a "polynucleotide" is a nucleic acid of more than one nucleotide. A polynucleotide can be made up of multiple polynucleotide units that are referred to by description of the unit. For example, a polynucleotide can comprise within its bounds a polynucleotide(s) having a coding sequence(s), a polynucleotide(s) that is a regulatory region(s) and/or other polynucleotide units commonly used in the art.

The present invention also relates to recombinant vectors and recombinant hosts, both prokaryotic and eukaryotic, which contain the substantially purified nucleic acid molecules disclosed throughout this specification. The DNA sequences of the present invention encoding a polypeptide disclosed herein, in whole or in part, can be linked with other DNA sequences, *i.e.*, a sequences to which the nucleic acid is not naturally linked, to form "recombinant DNA molecules" a nucleic acid disclosed herein. The novel DNA sequences of the present invention can be inserted into vectors in order to direct recombinant expression of polypeptides disclosed herein. Such vectors may be comprised of DNA or RNA; for most purposes DNA vectors are preferred. Typical vectors include plasmids, modified viruses, bacteriophage, cosmids, yeast artificial chromosomes, and other forms of episomal or

integrated DNA that can encode a polypeptide disclosed herein. One skilled in the art can readily determine an appropriate vector for a particular use.

An "expression vector" is a polynucleotide having regulatory regions operably linked to a coding region such that, when in a host cell, the regulatory regions can direct the expression of the coding sequence. The use of expression vectors is well known in the art. Expression vectors can be used in a variety of host cells and, therefore, the regulatory regions are preferably chosen as appropriate for the particular host cell. Preferred expression vectors can be those particularly designed for use in actinomycetes or the particular host chosen for a particular application of a gene or protein disclosed herein.

A "regulatory region" is a polynucleotide that can promote or enhance the initiation or termination of transcription or translation of a coding sequence. A regulatory region includes a sequence that is recognized by the RNA polymerase, ribosome, or associated transcription or translation initiation or termination factors of a host cell. Regulatory regions that direct the initiation of transcription or translation can direct constitutive or inducible expression of a coding sequence. Preferred regulatory regions can be those particularly designed for use in actinomycetes or the particular host chosen for a particular application of a gene or protein disclosed herein.

Polynucleotides of this invention contain full length or partial length sequences of ORFs disclosed herein. Polynucleotides of this invention can be single or double stranded. If single stranded, the polynucleotides can be a coding, "sense," strand or a complementary, "antisense," strand. Antisense strands can be useful as modulators of the receptor by interacting with RNA encoding the receptor. Antisense strands are preferably less than full length strands having sequences unique or highly specific for RNA encoding the receptor.

The polynucleotides can include deoxyribonucleotides, ribonucleotides or mixtures of both. The polynucleotides can be produced by cells, in cell-free biochemical reactions or through chemical synthesis. Non-natural or modified nucleotides, including inosine, methyl-cytosine, deaza-guanosine, and others known to those of skill in the art, can be present. Natural phosphodiester internucleotide linkages can be appropriate. However, polynucleotides can have non-natural linkages between the nucleotides. Non-natural linkages are well known in the art and include, without limitation, methylphosphonates, phosphorothioates, phosphorodithionates, phosphoroamidites and phosphate ester linkages. Dephospho-linkages are also

known, as bridges between nucleotides. Examples of these include siloxane, carbonate, carboxymethyl ester, acetamidate, carbamate, and thioether bridges. "Plastic DNA," having, for example, N-vinyl, methacryloxytethyl, methacrylamide or ethyleneimine internucleotide linkages, can be used. "Peptide Nucleic Acid" (PNA) is also useful and resists degradation by nucleases. These linkages can be mixed in a polynucleotide.

As used herein, "purified" and "isolated" are utilized interchangeably to stand for the proposition that the polynucleotides, proteins and polypeptides, or respective fragments thereof in question has been removed from its *in vivo* environment so that it can be manipulated by the skilled artisan, such as but not limited to sequencing, restriction digestion, site-directed mutagenesis, and subcloning into expression vectors for a nucleic acid fragment as well as obtaining the protein or protein fragment in quantities that afford the opportunity to generate polyclonal antibodies, monoclonal antibodies, amino acid sequencing, and peptide digestion. Therefore, the nucleic acids claimed herein can be present in whole cells or in cell lysates or in a partially, substantially or wholly purified form. A polynucleotide is considered purified when it is purified away from environmental contaminants. Thus, a polynucleotide isolated from cells is considered to be substantially purified when purified from cellular components by standard methods while a chemically synthesized nucleic acid sequence is considered to be substantially purified when purified from its chemical precursors.

Included in the present invention are nucleotide sequences that hybridize to the sequences disclosed herein under stringent conditions. By way of example, and not limitation, a procedure using conditions of high stringency is as follows: Prehybridization of filters containing DNA is carried out for 2 hr. to overnight at 65°C in buffer composed of 6X SSC, 5X Denhardt's solution, and 100 μg/ml denatured salmon sperm DNA. Filters are hybridized for 12 to 48 hrs at 65°C in prehybridization mixture containing 100 μg/ml denatured salmon sperm DNA and 5-20 X 10^6 cpm of ^{32}P-labeled probe. Washing of filters is done at 37°C for 1 hr in a solution containing 2X SSC, 0.1% SDS. This is followed by a wash in 0.1X SSC, 0.1% SDS at 50°C for 45 min. before autoradiography.

Other procedures using conditions of high stringency would include either a hybridization step carried out in 5XSSC, 5X Denhardt's solution, 50% formamide at 42°C for 12 to 48 hours or a washing step carried out in 0.2X SSPE, 0.2% SDS at 65°C for 30 to 60 minutes.

Reagents mentioned in the foregoing procedures for carrying out high stringency hybridization are well known in the art. Details of the composition of these reagents can be found in, *e.g.*, Sambrook, Fritsch, and Maniatis, 1989, Molecular Cloning: A Laboratory Manual, second edition, Cold Spring Harbor Laboratory Press. In addition to the foregoing, other conditions of high stringency which may be used are well known in the art.

Polypeptides

Preferred aspects of the present invention are substantially purified forms of the polypeptides encoded by the fragment of the *S. avermitilis* genome disclosed herein. Preferred embodiments of these aspects of the invention proteins that have an amino acid sequence which is set forth in SEQ ID NOs:2-10 and disclosed as follows in single letter code:

Peptide sequences:

ORF1 polypeptide

1 MAPDMNSQRF GGRLALVTGA GGGIGRATCA LGSAGARVVC VDRDGRGAGV
51 TPTWPERGAR AAWPEVADVS DGAAMERLPE RVAETYGVVD LLVNNAGIGM
101 AGRFLDTSVE DWQRTLGVNL WGVIHGCRLI GRQMAERGQG GHIVTVASAA
151 AFQPTRAVPA YATSKAAVLM LSECLRAEFA EFGVGVSVVC PGFVRTSFAS
201 AMHFAGVPRL EQERLRALFA GRGCSAEKVA AAVLRSVARD SAVVTVTAEA
251 RLSRLMSRFT PRLRAAVARM DPPS SEQ ID NO:2

ORF2 (AvrB) polypeptide

1 MSDHFLFMSA PFWGHVFPSL AVAEELVHRG HHVTFVTGAE MADAVRSVGA
51 DFLRYESAFE GVDMYRLMTE AEPNAIPMTL YDEGMSMLRS VEEHVGKDVP
101 DLVAYDIATS LNVGRVLAAS WSRPAMTVIP LFASNGRFST MQSVLDPDSA
151 QVSAPPPRFS EQMELFGLGA LVPRLAELLV SRGITEPVDD FLSGPEDFNL
201 VCLPRAFQYA GDTFDERFAF VGPCLGKRRG LGEWTPPGSG HPVVLISLGT
251 VFNRQLSFFR TFVRAFTDVP VHVVISLGKG VDPDVLRPLP PNVEVHRWVP
301 HHAVLEHARA LVTHGGTGSV MEALHAGCPV LVMPLSRDAQ VTGRRIAELG
351 LGRMVQPEEV TATTLRRHVL DIISDDAITR QVRQMQRATV EAGGALRAAD
401 ETERFLRRTR RH SEQ ID NO:3

ORF3b (AvrC) polypeptide

1 MRLLVTGGAG FIGSHFVRRL LTGAYPAFTG AEVVVLDKLT YAGRLENLAP
51 VLGSPSLIFV HGDICDGPLV ADLMDGSDMV VHFAAESHVD RSVADAAEFV
101 RTNVLGTHTL LRAATDAAVD RFVYISTDEV YGSIDSGSWT EDAPLEPNSP
151 YSASKASSDL LARSFHRTHG LPVIITRCSN NYGPHQFPEK LIPRFVTHLL
201 NGTKVPLYGD GENVRDWLHV DDHCRGIALV AERDRPGEIY HIGGGTELSN
251 RELTARLLDL LGVDWSMVEP VTDRKGHDRR YSLDISKISA ELGYAPRVPF
301 EEGLAQTVQW YVENRTLWEP LTARPELPVS DGASGAETAR SRPLPAGRRP
351 PRPWPAASA SEQ ID NO:4

ORF3a (AvrD) polypeptide

1 MKGIVLAGGT GSRLYPLTRA LSKQLLPVYD KPMIYYPLSV LMLGGIKDIL
51 VISSPDHLEQ FRRLLGDGSR LGLNIDYAAQ QRPGGIAEAF LIGADFIGQD
101 QVSLVLGDNI FHGMGFSHLL RSHTRDVDGC VLFGYAVTDP ERYGVGEVDA
151 SGKLLSVEEK PTAPRSNLAI TGLYLYDNDV IEVARGIRSS ARGELEITDV
201 NRAYLAEGRA RLVDLGRGFT WLDAGTHDSL MHAGQYVQVL EKRQGVRIAC
251 LEEIAFRMGL IDADDCYLRG VELAGSGYGE YLMSIAAEAA VRSPGCAYS SEQ ID NO:5

ORF4 (AvrE) polypeptide

1 MGRFSVCPPR PTGILKSMLT TGMCDRPLVV VLGASGYIGS AVAAELARWP
51 VLLRLVARRP GVVPPGGAAE TETRTADLTA ASEVALAVTD ADVVIHLVAR
101 LTQGAAWRAA ESDPVAERVN VGVMHDVVAA LRSGRRAGPP PVVVFAGSVY
151 QVGRPGRVDG SEPDEPVTAY ARQKLDAERT LKSATVEGVL RGISLRLPTV
201 YGAGPGPQGN GVVQAMVLRA LADEALTVWN GSVVERDLVH VEDVAQAFVS
251 CLAHADALAG RHWLLGSGRP VTVPHLFGAI AAGVSARTGR PAVPVTAVDP
301 PAMATAADFH GTVVDSSAFR AVTGWRPRLS LQEGLDHMVA AYV SEQ ID NO:6

ORF5 (AvrF) polypeptide

1 MTLGRPRRSS ADRPAPPAGA RATAAGVTVR RLVVEGAVEF TPTVFPDERG
51 LFVTPYQEPV LSEAVGHRFP TAQTCQSVSR RGVVRGVHFT ATPPGQAKYV
101 HCARGRALDF VVDLRTGSPT FGQWDSVLLD QERFRSVYLP IGVGHAFVAL
151 EDDTAMVYLM SSGYVPQNEH ALSPEDPDLA LPLGHHLGRA PILSERGPAR
201 APTLQQALRR GMLPEYRASR ALDEKL SEQ ID NO:7

ORF6 (AvrG) polypeptide

1 MPTTPSPAPL TARHDAALPA CLARSAAVGD TGARTSLDAF TGWWTRRSGA
51 HRFRVERIPF HGMDAWSFHP GTGNLAHRSG RFFSVEGLHV RGGEQPFPEW
101 QQPIIHQPEI GILGILAKKF DGVLHFLMQA KMEPGNINLV QLSPTVQATR
151 SNYTKVHGGA AVKYLEYFTQ PRRATVVVDV LQSEHGAWFH RKFNRNIVVE
201 TDEDVPLDDD FRWLTLGQIG ELMHRDNLVN MDARTVLACL PTPFDEPAAL
251 HSDAELLSWY AAERSRHSVH ARRVPLAGIP GWTTGAESIA HHADRYFRVV
301 AVRVEASNRE VAAWTQPLIE PCGHGITAFL TRRIGGVPHL LAHGRVEGGF
351 LDTIELGPTV QYTPRNYAHL TGPARPRFLD LVLEAAPDRI RYAAVHSEEG
401 GRFLHAQARY LFVEADESQA PNDPPPGYRW CTPGQLTQLL RYGRYVNVQA
451 RTLLSLLTTR AVEL SEQ ID NO:8

ORF7 (AvrH) polypeptide

1 MTEREFTDPR IVPHESEQER AAREQLTKLL VDSPIPPKYL IDNLSVYMRR
51 NQLADLLSMD ALYRMLPEVP GVIMEFGVLH GRHLATLTAL RSIYEPYNSL
101 RRVIGFDTFT GFPDIDEADE VSTSAVPGRF AVPDGEVEHL RQVLAAHEAN
151 EPYGHTQRSF VVQGDVRETV PQYLAEHPHT VIALAYFDLD LYRPTRELLD
201 VITPHLTRGS ILAFDELTHP KWPGETRALS EAFGLDHAPL RQLPGREPPV
251 IYMTWGA SEQ ID NO:9

ORF8 (AvrI) polypeptide

1 MMPTTAEPAP VQSDSNSSAP LHTELGRTRL RISRLALGTV NIGGRVEEPE
51 ARRLMDHALA QGITLFDTAN TYGWRVHKGY TEEVIGRWLA DRPARREQVV
101 LATKVGDPMG SGPNDHGLSV RNIVAACDAS LRRLRTDWID LYQLHHIDRR
151 AGWDEVWQAM DLLITQGKVR YVGSSNFAGW DIASAQEAAR RRNALGLASE
201 QCVYNLVTRH AELEVIPAAS AYGVGVLVWS PLHGGLLGGV LRKTRENTAV
251 KSAQGRAVEA LEHHRTTIAA YEDVCADHGL DPAHVGMAWV LSRPGVTGLV
301 IGPRTEQHVD GALHALRTPL PEPVLARLEE LFPPVGRGGS APDAWLS SEQ ID NO:10

The present invention also relates to fragments and mutant or polymorphic forms of the proteins set forth in SEQ ID NOs:2-10, including but not necessarily limited to amino acid substitutions, deletions, additions, amino terminal truncations and carboxy-terminal truncations such that these provide for proteins or protein fragments of enzymatic, biocatalytic, biosynthetic or diagnostic use.

Using the disclosure of polynucleotide and polypeptide sequences provided herein to isolate polynucleotides encoding naturally occurring forms of the proteins disclosed herein, one of skill in the art can determine whether such naturally occurring forms are mutant or polymorphic forms by sequence comparison. One can determine whether the mutant or polymorphic forms, or fragments of any protein disclosed herein, are biologically active by routine testing of the protein or fragment in a *in vitro* or *in vivo* assay for the biological activity of the full length version of the protein as encoded by the nucleotide sequence disclosed herein. For example, one can express N-terminal or C-terminal truncations, or internal additions or deletions of a protein in a host cell and test whether the altered form can perform the same enzymatic step as performed by the full-length polypeptide disclosed herein.

It is known that there is a substantial amount of redundancy in the various codons which code for specific amino acids. Therefore, this invention is also directed to those DNA sequences encode RNA comprising alternative codons which code for the eventual translation of the identical amino acid sequence of any of the avermectin pathway proteins disclosed herein. Therefore, the present invention includes nucleic acid sequences that vary because of codon redundancy which can result in differing DNA molecules expressing an identical protein.

As with many enzymes, it is possible to modify many of the amino acids of the proteins disclosed herein, particularly those which are not found in the ligand binding or catalytic domains, and still retain substantially the same biological activity as the original protein. Thus this invention includes modified polypeptides which have amino acid deletions, additions, or substitutions but that still retain substantially the same biological activity as proteins disclosed herein. Also included within the scope of this invention are polypeptides having changes which do not substantially alter the ultimate physical or functional properties of the expressed protein. A "conservative amino acid substitution" refers to the replacement of one amino acid residue by another, chemically similar, amino acid residue. Examples of such conservative substitutions are: substitution of one hydrophobic residue (isoleucine, leucine, valine, or methionine) for another; substitution of one polar residue for another polar residue of the same charge (*e.g.*, arginine for lysine; glutamic acid for aspartic acid). In particular, substitution of valine for leucine, arginine for lysine, or asparagine for glutamine is not expected to cause a change in functionality of the polypeptide.

It is known that DNA sequences coding for a peptide can be altered so as to code for a peptide having properties that are different than those of the naturally occurring peptide. Methods of altering the DNA sequences include but are not limited to site directed mutagenesis. Examples of altered properties include but are not limited to changes in the affinity of an enzyme for a substrate or a receptor for a ligand.

For the purposes of this invention, naturally occurring, or wild-type protein has an amino acid sequence shown as SEQ ID NOs:2-10 and is encoded by the particular nucleic acid sequences disclosed herein. As used herein, a "functional equivalent" of a wild-type protein possesses a biological activity that is substantially the same biological activity of the wild type protein. A polypeptide has "substantially the same biological activity" as a wild-type if that polypeptide has a K_d for a ligand that is no more than 5-fold greater than the K_d of the wild-type for the same ligand. The term "functional derivative" is intended to include those "fragments," "mutants," "variants," "degenerate variants," "analogs," "homologues" or "chemical derivatives" of the wild type protein that exhibit substantially the same biological activity. The term "fragment" is meant to refer to any polypeptide subset of wild-type protein disclosed herein. The term "mutant" is meant to refer to a molecule that may be substantially similar to the wild-type form but possesses distinguishing biological characteristics. Such altered characteristics include but are in no way limited to altered substrate binding, altered substrate affinity and altered sensitivity to chemical compounds affecting biological activity of the wild-type. The term "variant" is meant to refer to a molecule substantially similar in structure and function to either the entire wild-type protein or to a fragment thereof.

As used herein in reference to a gene or encoded protein, a "polymorphic" form that is naturally found as an allele in the population at large. A polymorphic form can have a different nucleotide sequence from the particular nucleic acid or protein disclosed herein. However, because of silent mutations, a polymorphic gene can encode the same or different amino acid sequence as that disclosed herein. Further, some polymorphic forms will exhibit biological characteristics that distinguish the form from wild-type protein activity, in which case the polymorphic form is also a mutant. Polymorphic forms encompass allelic variants.

A protein or fragment thereof is considered purified or isolated when it is obtained at a concentration at least about five-fold to ten-fold higher than that found

in nature. A protein or fragment thereof is considered substantially pure if it is obtained at a concentration of at least about 100-fold higher than that found in nature. A protein or fragment thereof is considered essentially pure if it is obtained at a concentration of at least about 1000-fold higher than that found in nature.

Expression of Proteins of this Invention

The present invention also relates to recombinant vectors and recombinant hosts, both prokaryotic and eukaryotic, which contain the substantially purified nucleic acid molecules disclosed throughout this specification.

Therefore, the present invention also relates to methods of expressing the proteins and their biological equivalents described herein and reactions employing these recombinantly expressed gene products, including *in vivo* or *in vitro* biosynthetic, biocatalytic or biotransformation reactions employing the genes, proteins, vectors and host cells disclosed herein.

A variety of expression vectors can be used to express recombinant proteins in host cells. Expression vectors are defined herein as DNA sequences that are arranged for the transcription of cloned DNA and the translation of their mRNAs in an appropriate host. Such vectors can be used to express the nucleotide sequences of this invention in a variety of hosts such as bacteria, blue-green algae, plant cells, insect cells and animal cells. Specifically designed vectors allow the shuttling of DNA between hosts such as bacteria-yeast or bacteria-animal cells. An appropriately constructed expression vector should contain: an origin of replication for autonomous replication in host cells, selectable markers, a limited number of useful restriction enzyme sites, optionally a potential for high copy number, and promoters. A promoter is defined as a DNA sequence operably linked to a coding region so that it interacts with cellular proteins to direct RNA polymerase to bind to DNA and initiate mRNA synthesis. A strong promoter is one which causes mRNAs to be initiated at high frequency. A promoter can be inducible. Expression vectors can include, but are not limited to, cloning vectors, modified cloning vectors, specifically designed plasmids or viruses.

Commercially available mammalian expression vectors which can be suitable for recombinant protein expression, include but are not limited to, pcDNA3.1 (Invitrogen), pLITMUS28, pLITMUS29, pLITMUS38 and pLITMUS39 (New England Biolabs), pcDNAI, pcDNAIamp (Invitrogen), pcDNA3 (Invitrogen), pMC1neo (Stratagene), pXT1 (Stratagene), pSG5 (Stratagene), EBO-pSV2-neo

(ATCC 37593) pBPV-1(8-2) (ATCC 37110), pdBPV-MMTneo(342-12) (ATCC 37224), pRSVgpt (ATCC 37199), pRSVneo (ATCC 37198), pSV2-dhfr (ATCC 37146), pUCTag (ATCC 37460), and lZD35 (ATCC 37565).

A variety of bacterial expression vectors can be used to express recombinant protein in bacterial cells. Commercially available bacterial expression vectors which are suitable for recombinant expression include, but are not limited to pQE (Qiagen), pET11a (Novagen), lambda gt11 (Invitrogen), and pKK223-3 (Pharmacia). Preferred vectors include vectors designed for expression of proteins in actinomycetes including but not limited to the pIJ series developed at the John Innes Institute and described in Hopwood, D.A. *et al.*, 1985. Genetic Manipulation of Streptomyces, A Laboratory Manual. F. Crowe & Sons, Ltd., Norwich, England.) A variety of fungal cell expression vectors can be used to express recombinant protein in fungal cells. Commercially available fungal cell expression vectors which are suitable for recombinant expression include but are not limited to pYES2 (Invitrogen) and *Pichia* expression vector (Invitrogen).

A variety of insect cell expression vectors can be used to express recombinant receptor in insect cells. Commercially available insect cell expression vectors which are suitable for recombinant expression include but are not limited to pBlueBacIII and pBlueBacHis2 (Invitrogen), and pAcG2T (Pharmingen).

An expression vector containing DNA encoding a protein can be used for expression of the protein in a recombinant host cell. Recombinant host cells can be prokaryotic or eukaryotic, including but not limited to bacteria such as *E. coli* or *Streptomycetes*, fungal cells such as yeast, mammalian cells including but not limited to cell lines of human, bovine, porcine, monkey and rodent origin, and insect cells including but not limited to Drosophila- and silkworm-derived cell lines. Cell lines derived from mammalian species which can be suitable and which are commercially available, include but are not limited to, L cells L-M(TK$^-$) (ATCC CCL 1.3), L cells L-M (ATCC CCL 1.2), Saos-2 (ATCC HTB-85), 293 (ATCC CRL 1573), Raji (ATCC CCL 86), CV-1 (ATCC CCL 70), COS-1 (ATCC CRL 1650), COS-7 (ATCC CRL 1651), CHO-K1 (ATCC CCL 61), 3T3 (ATCC CCL 92), NIH/3T3 (ATCC CRL 1658), HeLa (ATCC CCL 2), C127I (ATCC CRL 1616), BS-C-1 (ATCC CCL 26), MRC-5 (ATCC CCL 171) and CPAE (ATCC CCL 209). The appropriateness of any cell line for any particular purpose can be assessed by simply testing the expression of a protein of this invention in the cell line.

The expression vector can be introduced into host cells via any one of a number of techniques including but not limited to transformation, transfection, protoplast fusion, and electroporation. The expression vector-containing cells are analyzed to determine whether they produce protein. Identification of expressing cells can be done by several means, including but not limited to immunological reactivity with antibodies, labeled ligand binding and the presence of host cell-associated recombinant protein activity.

The cloned DNA obtained through the methods described herein can be recombinantly expressed by molecular cloning into an expression vector containing a suitable promoter and other appropriate transcription regulatory elements, and transferred into prokaryotic or eukaryotic host cells to produce recombinant protein. Techniques for such manipulations can be found described in Sambrook, et al., *supra* , and are well known and easily available to the one of ordinary skill in the art.

Expression of protein can also be performed using *in vitro* produced synthetic mRNA. Synthetic mRNA can be efficiently translated in various cell-free systems, including but not limited to wheat germ extracts and reticulocyte extracts.

To determine the sequence(s) that yields optimal levels of recombinant protein, molecules including but not limited to the following can be constructed: a DNA fragment containing the full-length open reading frame for a protein as well as various constructs containing portions of the DNA encoding only specific domains of the protein or rearranged domains of the protein. The expression levels and activity of the protein can be determined following the introduction, both singly and in combination, of these constructs into appropriate host cells. Following determination of the DNA cassette yielding optimal expression in transient assays, this construct is transferred to a variety of expression vectors, including but not limited to those for mammalian cells, plant cells, insect cells, oocytes, bacteria, and yeast cells where expression is assessed.

Following expression of a recombinant protein in a host cell, the recombinant polypeptides can be recovered. Several protein purification procedures are available and suitable for use. Protein and polypeptides can be purified from cell lysates and extracts, or from conditioned culture medium, by various combinations of, or individual application of methods including ultrafiltration, acid extraction, alcohol precipitation, salt fractionation, ionic exchange chromatography, phosphocellulose chromatography, lecithin chromatography, affinity (*e.g.*, antibody or His-Ni) chromatography, size exclusion chromatography, hydroxylapatite adsorption

chromatography and chromatography based on hydrophobic or hydrophilic interactions. In some instances, protein denaturation and refolding steps can be employed. High performance liquid chromatography (HPLC) and reversed phase HPLC can also be useful. Dialysis can be used to adjust the final buffer composition.

Antibodies

The present invention also relates to polyclonal and monoclonal antibodies raised in response to a protein disclosed herein, or a fragment thereof. It is preferable to raise antibodies to epitopes which show the least homology to other known proteins.

An antibody is specific for an epitope if one of skill in the art can use standard techniques to determine conditions under which one can detect a polypeptide of this invention in a Western Blot of a sample from a host cell that expresses a protein of this invention. The blot can be of a native or denaturing gel as appropriate for the epitope. An antibody is highly specific for an epitope if no nonspecific background binding is visually detectable. An antibody can also be considered highly specific if the binding of the antibody to the protein can not be competed by non-homologous peptides, polypeptides or proteins, but can be competed by homologous peptides or polypeptides or the full length form of the relevant protein as disclosed herein.

Recombinant protein can be separated from other cellular proteins by use of an immunoaffinity column made with monoclonal or polyclonal antibodies specific for full-length protein, or polypeptide fragments of protein. Additionally, polyclonal or monoclonal antibodies can be raised against a synthetic peptide (usually from about 9 to about 25 amino acids in length) from a portion of a protein disclosed in SEQ ID NOs:2-10. Monospecific antibodies are purified from mammalian antisera containing antibodies reactive against a protein or are prepared as monoclonal antibodies reactive with a protein using the technique of Kohler and Milstein (1975, *Nature* 256: 495-497). Monospecific antibody as used herein is defined as a single antibody species or multiple antibody species with homogenous binding characteristics for a particular protein. Homogenous binding as used herein refers to the ability of the antibody species to bind to a specific antigen or epitope, such as those associated with a protein described herein. Specific antibodies are raised by immunizing animals such as mice, rats, guinea pigs, rabbits, goats, horses and the like, with an appropriate concentration of a protein described herein or a synthetic

peptide generated from a portion of the described proteins with or without an immune adjuvant.

Preimmune serum is collected prior to the first immunization. Each animal receives between about 0.1 mg and about 1000 mg of protein associated with an acceptable immune adjuvant. Such acceptable adjuvants include, but are not limited to, Freund's complete, Freund's incomplete, alum-precipitate, water in oil emulsion containing *Corynebacterium parvum* and tRNA. The initial immunization consists of injecting protein or peptide fragment thereof, preferably in Freund's complete adjuvant, at multiple sites either subcutaneously (SC), intraperitoneally (IP) or both. Each animal is bled at regular intervals, preferably weekly, to determine antibody titer. The animals may or may not receive booster injections following the initial immunization. Those animals receiving booster injections are generally given an equal amount of protein in Freund's incomplete adjuvant by the same route. Booster injections are given at about three week intervals until maximal titers are obtained. At about 7 days after each booster immunization or about weekly after a single immunization, the animals are bled, the serum collected, and aliquots are stored at about -20°C.

Monoclonal antibodies (mAb) reactive with a protein are prepared by immunizing inbred mice, preferably Balb/c, with the protein. The mice are immunized by the IP or SC route with about 1 mg to about 100 mg, preferably about 10 mg, of protein in about 0.5 ml buffer or saline incorporated in an equal volume of an acceptable adjuvant, as discussed herein. Freund's complete adjuvant is preferred. The mice receive an initial immunization on day 0 and are rested for about 3 to about 30 weeks. Immunized mice are given one or more booster immunizations of about 1 to about 100 mg of protein in a buffer solution such as phosphate buffered saline by the intravenous (IV) route. Lymphocytes, from antibody positive mice, preferably splenic lymphocytes, are obtained by removing spleens from immunized mice by standard procedures known in the art. Hybridoma cells are produced by mixing the splenic lymphocytes with an appropriate fusion partner, preferably myeloma cells, under conditions which will allow the formation of stable hybridomas.

Fusion partners can include, but are not limited to: mouse myelomas P3/NS1/Ag 4-1; MPC-11; S-194 and Sp 2/0, with Sp 2/0 being preferred. The antibody producing cells and myeloma cells are fused in polyethylene glycol, about 1000 mol. wt., at concentrations from about 30% to about 50%. Fused hybridoma cells are selected by growth in hypoxanthine, thymidine and aminopterin

supplemented Dulbecco's Modified Eagles Medium (DMEM) by procedures known in the art. Supernatant fluids are collected form growth positive wells on about days 14, 18, and 21 and are screened for antibody production by an immunoassay such as solid phase immunoradioassay (SPIRA) using the protein as the antigen. The culture fluids are also tested in the Ouchterlony precipitation assay to determine the isotype of the mAb. Hybridoma cells from antibody positive wells are cloned by a technique such as the soft agar technique of MacPherson, 1973, Soft Agar Techniques, in *Tissue Culture Methods and Applications*, Kruse and Paterson, Eds., Academic Press.

Monoclonal antibodies are produced *in vivo* by injection of pristine primed Balb/c mice, approximately 0.5 ml per mouse, with about 2 x 10^6 to about 6 x 10^6 hybridoma cells about 4 days after priming. Ascites fluid is collected at approximately 8-12 days after cell transfer and the monoclonal antibodies are purified by techniques known in the art.

In vitro production of mAb is carried out by growing the hybridoma in DMEM containing about 2% fetal calf serum to obtain sufficient quantities of the specific mAb. The mAb are purified by techniques known in the art.

Antibody titers of ascites or hybridoma culture fluids are determined by various serological or immunological assays which include, but are not limited to, precipitation, passive agglutination, enzyme-linked immunosorbent antibody (ELISA) technique and radioimmunoassay (RIA) techniques. Similar assays are used to detect the presence of the protein in a biological sample or in an *in vitro* biocatalysis reaction.

It is readily apparent to those skilled in the art that the herein described methods for producing monospecific antibodies can be utilized to produce antibodies specific for peptide fragments, or full-length proteins described herein.

Antibody affinity columns are made, for example, by adding the antibodies to Affigel-10 (Biorad), a gel support which is pre-activated with N-hydroxysuccinimide esters such that the antibodies form covalent linkages with the agarose gel bead support. The antibodies are then coupled to the gel via amide bonds with the spacer arm. The remaining activated esters are then quenched with 1M ethanolamine HCl (pH 8). The column is washed with water followed by 0.23 M glycine HCl (pH 2.6) to remove any non-conjugated antibody or extraneous protein. The column is then equilibrated in phosphate buffered saline (pH 7.3) and the cell culture supernatants or cell extracts containing full-length protein or protein fragments are slowly passed through the column. The column is then washed with phosphate

buffered saline until the optical density (A_{280}) falls to background, then the protein is eluted with 0.23 M glycine-HCl (pH 2.6). The purified protein is then dialyzed against phosphate buffered saline.

Levels of recombinant protein in host cells is quantified by a variety of techniques including, but not limited to, immunoaffinity and/or ligand affinity techniques. Specific-antibody affinity beads or specific antibodies are used to isolate ^{35}S-methionine labeled or unlabelled recombinant protein. Labeled recombinant protein is analyzed by SDS-PAGE. Unlabelled protein is detected by Western blotting, ELISA or RIA assays employing either protein specific antibodies and/or antiphosphotyrosine antibodies.

Avermectin Glycosylation Genes and Proteins

A cluster of genes involved in the synthesis and/or addition of oleandrose to avermectin aglycone has been cloned. pVE650, a 47.8 kb plasmid was isolated from a library of *S. avermitilis* by its ability to complement a mutant producing non-glycosylated avermectins. Five overlapping cosmid clones of *S. avermitilis* genomic DNA were isolated using a fragment of pVE650 as a probe. Subclones from pVE650 and an overlapping cosmid were used in complementation studies with 23 mutants defective in the glycosylation of avermectin aglycone. Seven complementation classes were identified. A 11-kb *PstI* fragment of *S. avermitilis* genomic DNA complemented all 23 mutants, indicating the genes for avermectin glycosylation were clustered. The 11 kb PstI fragment can be cloned from a deposited strain, ATCC 67890, which contains plasmid pVE859

The 11 kb PstI fragment of the avermectin gene cluster from *S. avermitilis* was subcloned into an integration vector, pVE1053. The resulting plasmid, pVE1190 could complement all the mutants known to us that are defective in glycosylation. The result indicated that pVE1190 encoded all the genes for biosynthesis and attachment of oleandrose disaccharide to avermectin aglycone. Upon sequencing 10 kb region of the fragment, it was discovered that the fragment contained nine open reading frames.

The 11-kb subclone was mutagenized with Tn*5* and Tn*5*seq1. Fourteen insertions were transferred to *S. avermitilis* and used in complementation analysis. An eighth complementation class was identified. Sequencing of an 10-kb region identified nine ORFs and an additional partial ORF. Eight of the nine ORFs were correlated to seven glycosylation complementation classes confirming that these

eight genes are involved in the biosynthesis and attachment of oleandrose to avermectin aglycones. Sequence comparison to Genbank data bases identified 6 of the genes as: dTDP-glucase synthase(ORF 3a), dTDP-glucose 4,6 dehydrase(ORF3b), dTDP-4-keto-hexose reductase (ORF4), dTDP-hexose 3,5 epimerase(ORF5), dTDP-hexose 3' O-methylase(ORF7), and an avermectin aglycone-dTDP-oleandrose glycosyltransferase(ORF2). The ninth ORF was essential for biosynthesis of the avermectin aglycones. The partial ORF encoded part of an avermectin polyketide synthase module 7.

The genes from this cluster or the encoded polypeptides could be used to glycosylate avermectin aglycones or other macrolide aglycones. For instance US patent US 5,312,753 describes the glucosylation of the C13 and C14a positions of avermectin derivatives by a *S. avermitilis* strain. Another use of the polynucleotides and polypeptides would be to use them separately and in combination with other cloned genes or expressed proteins to make and attach known and novel sugars to known and novel macrolides or to other hydroxyl containing compounds.

EXAMPLE 1

Cloning of the Gene Cluster

Bacterial strains and plasmids.

Ligation mixtures were used to transform E. coli MM294 (*E. coli* Genetic Stock Center, New Haven, CT.) Derivatives of pVE616 were isolated from the triply DNA methylase deficient host ET12567. The isolation of mutants deficient in glycosylation of avermectin aglycones has been described (Ruby, *et al.*, 1990). Some of the glycosylation mutants were isolated from a mutant deficient in C-5 O-methylation of avermectin (Ruby 1986, Ruby *et al.*, 1990). *S. avermitilis* mutants deficient in 3',3" O-methylation (GMT) have been described (Ruby et a., 1985). *S. lividans* strain 1326 and its SLP2$^-$SLP3$^-$ derivative TK21 (Hopwood *et al.*, 1983) were obtained from D. Hopwood (John Innes Institute, Norwich, UK). pBR322 was obtained from BRL (Bethesda, MD) and pIJ922 was obtained from D. Hopwood (Hopwood *et al.*, 1985). pVE616 is a 4.4 kb Amp^R derivative of pBR322 which contains a 1.8 kb *Bam*HI fragment which expresses thiostrepton-resistance in *Streptomyces* (Gene). Cultures were preserved by adding 0.1 ml of dimethyl

sulfoxide (Aldrich Chemical Co., Milwaukee, WI) to 0.9 ml of culture and quick freezing the mixture at -70°C.

Media, Solutions, and Chemicals

Streptomyces were grown as dispersed cultures for the isolation of chromosomal or plasmid DNA in YEME medium (Thompson, *et al.*, 1982) with 30% sucrose and 0.25% glycine. *E. coli* was grown in LB (Miller, 1972). Solid media containing 1.5% agar included LB for *E. coli* (Miller, 1972), R2YE for *S. lividans* (Thompson, *et al.*, 1982), RM14 for *S. avermitilis* (MacNeil & Klapko, 1987), and YME-TE for *S. avermitilis* . YME-TE contained per liter: yeast extract 3.0 g, malt extract 10.0 g, dextrose 4.0 g and 4 ml of a trace element solution (per liter: HCl (37.3%) 49.7 ml, $MgSO_4 \cdot 7H_2O$ 61.1g, $CaCO_3$ 2.0g, $FeCl_3 \cdot 6H_2O$ 5.4 g, $ZnSO_4 \cdot 7H_2O$ 1.44 g, $MnSO_4 \cdot H_2O$ 1.11 g, $CuSO_4 \cdot 5H_2O$ 0.25 g, H_3BO_3 0.062g, $Na_2MoO_4 \cdot 2H_2O$ 0.49 g). YME-TE was adjusted to pH 7.0 with NaOH before autoclaving. Fermentation medium A, contained, per liter: glucose 20.0 g, yeast extract 20.0g, Hy-Case SF 20.0 g/ml, $MgSO_4 \cdot 7H_2O$ (12.5%), NaCl (12.5%), $MnSO_4 \cdot H_2O$ (0.5%), $ZnSO_4 \cdot 7H_2O$ (1.0%), $CaCl_2 \cdot 2H_2O$ (2.0%), $FeSO_4 \cdot 7H_2O$ 0.025 g, and KNO_3 2.0 g. Fermentation medium B, which was adjusted to pH 7.2 with NaOH before autoclaving contained, per liter, peptonized milk 20.0 g, Ardamine pH 4.0 g, glucose 90.0 g, $MgSO_4 \cdot 7H_2O$ 0.5 g, $CuSO_4 \cdot 5H_2O$ (0.06 mg/ml) 1 ml, $ZnSO_4 \cdot 6H_2O$ (1 mg/ml) 1 ml, $CoCl_2 \cdot 6H_2O$ (0.1 mg/ml) 1ml, and $FeCl_2 \cdot 6H_2O$ (3 mg/ml) 1 ml. TE buffer (10 mM Tris, pH 7.9, 1 mM EDTA) was used to store and dilute DNA. Polyethylene glycol 1000 (PEG), agarose and ampicillin were obtained from Sigma Chemical Co., St Louis, MO. Formamide was obtained from IBI (New Haven, CT). Thiostrepton (gift from E. R. Squibb & Sons, Princeton, NJ) was added to a final concentration of 5 μg/ml in liquid medium, 10 μg/ml in solid medium, and 15 μg/ml when added as an overlay to select transformants. Ampicillin was added to a final concentration of 100 μg/ml.

Isolation of DNA

Large (500 ml) and small (1.5 ml) scale preparations of plasmid DNA were isolated from *E. coli* by the alkaline lysis procedure (Maniatis *et al.* 1982). A modified alkaline lysis procedure was developed for *Streptomyces*. Small scale plasmid preparations were prepared form cultures grown in 5 ml of YEME and

washed as described previously (MacNeil, 1987). Cell pellets were resuspended in 1 ml of 50 mM glucose, 25 mM Tris pH 8, 10 mM EDTA, and 50 μl of a 15 mg/ml lysozyme solution in 50 mM glucose, 25 mM Tris pH 8, 10 mM EDTA was added. Following incubation for 15 minutes at 37°C, 1.5 ml of a 0.2 N NaOH, 1% SDS solution was added, the mixture was vortexed for 5 seconds and the mixture was incubated for 15 minutes on ice. Next 150 μl of ice cold pH 4.8 potassium acetate solution (5 M with respect to acetate, 3 M with respect to potassium) was added, the mixture vortexed for 10 seconds, and incubated on ice for 15 minutes. The mixture was centrifuged for 15 minutes at 12,000 x g, at 4°C and the resulting supernatant was transferred to a new tube. 2.0 ml of -20°C isopropanol or isopropanol containing 0.05% diethyl pyrocarbonate was added, mixed, and centrifuged at 12,000 x g for 15 minutes at 4°C. The DNA pellet was dried and the DNA was dissolved in 0.5 ml of 0.3 M ammonium acetate. The solution was transferred to a 1.5 ml Eppendorf tube, mixed with 400 μl of phenol, previously equilibrated with 1 M Tris pH 7.9, and the aqueous phase separated by centrifugation in a microfuge for 3 minutes. The aqueous phase was removed to another Eppendorf tube and extracted with 400 μl of chloroform. The resulting aqueous DNA solution was precipitated with 2 volumes of ethanol, washed with 70% ethanol, and the plasmid DNA resuspended in 100 μl of TE. Large scale plasmid preparations were isolated from 1 l YEME cultures of *Streptomyces* by a scaled up alkaline lysis procedure except that the DNA precipitated by isopropanol was resuspended in a CsCl solution and subjected to two bandings. Chromosomal DNA from *Streptomyces* was prepared as described by Hopwood *et al.*, 1985.

Transformations with plasmid DNA.

The procedures for preparation of protoplasts, storage of protoplasts, polyethylene glycol mediated transformation of protoplasts, regeneration of protoplasts, and selection of transformants has been described for *S. lividans* (MacNeil, 1987) and *S. avermitilis* (MacNeil & Klapko, 1987). Transformation of *E. coli* with plasmid DNA has been described (Maniatis *et al.*, 1982).

Restriction enzyme analysis

Restriction enzymes were obtained from New England Biolabs (Beverly, MA), Bethesda Research Labs (Bethesda, MD), or IBI (New Haven, CT)

and were used according to the manufactures directions. Agarose gels were prepared and electrophoresis performed as described (Maniatis *et al.*, 1982).

Construction of subclones from pVE650 and pVE859

Restriction fragments to be used in the construction of subclones from pVE650 and pVE859 were purified from agarose gel slices by electroelution and ligated to CIAP treated vector DNA. Subclones into pVE616 were transformed into MM294 and the appropriate constructs were identified. Plasmid DNA was transformed into ET12567 (a triply DNA methylation deficient strain), purified by CsCl centrifugation, and 5 μg of the resulting DNA was used to transform *S. avermitilis* . Subclones into pIJ922 were transformed into *S. lividans* TK21, analyzed, purified from CsCl gradients, and 100ng of the plasmid DNA was transformed into *S. avermitilis* .

S. avermitilis fermentations and analysis of avermectin production

Single colonies from transformation plates were picked with a sterile toothpick on to YME-TE medium and subjected to small scale solid fermentations as described MacNeil et al., 1992. After 12-16 days incubation at 27-28°C, the mycelia was extracted with methanol, aliquots of the extract were applied to E. Merck Silica Gel 60 F-254 TLC plates and the avermectins developed for 15 minutes with a dichloromethane: ethylacetate:methanol 9:9:1 solvent mixture. Avermectins are visualized under UV illumination. Under these conditions 4 glycosylated avermectins are resolved from strains which produce wild type avermectins. OMT- cultures produce predominantly the B avermectins. Mutants unable to glycosylate avermectin aglycones also produce 4 bands, however, since aglycones are a better substrate for the C5-Omethyltransferae, mostly the A- aglycones are produced. In contrast in the OMT- strains, residual C5-O-methyltransferase only methylates about 1/2 the aglycones resulting in 4 bands. The aglycones run faster in the TLC system than the corresponding glycosylated avermectins. As shown previously (Gene) the order, from fastest to slowest band is, avermectin aglycone A_1a+b, avermectin aglycone A_2a+b, avermectin A_1a+b and avermectin aglycone B_1a+b, avermectin A_2a+b and avermectin aglycone B_2a+b, avermectin B_1a+b, and avermectin B_2a+b.

Colony hybridizations

The cosmid library of *S. avermitilis* was constructed in the 6.7 kb, double lambda *cos* vector, pVE328, and consists of 2016 cosmid clones stored as individual cultures in 21 microtiter dishes. Replicates of the library were made on LB plates containing ampicillin, colonies were transferred to Biotrans nylon membranes (1.5 μM pore size), and colonies processed to release and fix DNA to the filters (Maniatis *et al.*,). The resulting 21 filters were individually hybridized with ^{32}P labeled probes. Preparation of probes, hybridizations and autoradiography were as described above for Southern analysis. Putative hybridizing clones were retested by patching duplicates to LB plates with ampicillin, lifting the colonies to nitrocellulose (Schleicher & Schuell, Keene, NH), fixing the DNA to the filters and hybridizing with the probe. Plasmid DNA was isolated from the cosmid clones which retested positive, restricted with *Bam*HI, and confirmed by a Southern analysis.

Isolation of pVE650

S. avermitilis produces 8 major avermectins which can be separated by TLC into 4 bands representing, from most polar to least, avermectin A_1a+b, A_2a+b, B_1a+b and B_2a+b. A pIJ922 based library of *S. avermitilis* DNA was constructed and screened for complementation of two mutants defective in avermectin biosynthesis (Avr). One mutant was a C-5 O-methyltransferase mutant (OMT), which produces predominantly avermectin B_1a+b and B_2a+b. The other mutant was MA6278, an avermectin aglycone producer. Several overlapping plasmids were isolated which complemented OMT mutants (Streicher *et al*). When the plasmids which complemented OMT mutants were introduced into several mutants altered in, or defective in, avermectin biosynthesis, no other mutants were complemented (Streicher *et al*). Approximately 3000 transformants of MA6278 were screened for avermectin production by small scale fermentation and TLC analysis of methanol extracts of each transformant. One transformant complemented the defect in MA6278. A plasmid was isolated from this transformant and designated pVE650. The presence of avermectin glycosylation genes on pVE650 was confirmed by retransforming MA6278 by pVE650 and detecting glycosylated avermectins by TLC. Most aglycone producing mutants (21/26) were complemented by pVE650.

Physical analysis of pVE650

A restriction map was determined for pVE650 see MacNeil et al, 1992. The insert in pVE650 is delimited by *Bam*HI sites, no sites were found in the 24 kb

insert for the following enzymes: *Ase*I, *Dra*I, *Eco*RV, *Hind*III, *Hpa*I, *Nde*I, *Nhe*I, *Spe*I, *Ssp*I, and *Xba*I. No common restriction bands were found between pVE650 and pAT1, a plasmid which complements OMT mutants (Streicher *et al*).

The insert in pVE650 was found to be colinear with the chromosome of *S. avermitilis* by Southern analysis. The 9 *Bam*HI fragments greater than 400 bp were used as probes against *Bam*HI and *Sst*I digestions of genomic DNA from avermectin producing and nonproducing strains. Seven of the nine *Bam*HI fragments hybridized to a band identical in size to the *Bam*HI fragment used as probes. Therefore, the seven *Bam*HI fragments do not appear to have undergone rearrangement to form pVE650. This was confirmed by the *Sst*I digestions in which adjacent *Bam*HI fragments hybridize to an overlapping *Sst*I fragment. Two *Bam*HI fragments at the ends of the insert in pVE650, the 2.1 kb and 1.1 kb fragments, hybridized to larger fragments. These results indicate that pVE650 resulted from the ligation of a *Sau*3AI fragment into the *Bam*HI site of pIJ922 in such a way that *Bam*HI sites formed at both junctions.

EXAMPLE 2

Identification of the Genes for Avermectin Glycosylation

Identification of three genes for avermectin glycosylation on pVE650

The 26 AGL$^-$ mutants were divided into 4 complementation classes by introducing subclones of pVE650 into the AGL$^-$ mutants. Complementation tests were performed by introducing subclones into various aglycone producing mutants and testing transformants for the ability to produce avermectins or avermectin aglycones. Fragments form pVE650 were subcloned into pIJ922, a low copy number *Streptomyces* vector (Hopwood *et al.*, 1985), or pVE616, an *E. coli* vector that fails to replicate in *Streptomyces* but which can integrate by recombination between the chromosome and the cloned fragment. Between 6 and 12 transformants were tested for avermectin production as visualized by TLC analysis of fermentation extracts. Occasionally, an individual transformant failed to produce avermectin aglycones or avermectins. Positive complementation was scored if at least 5/6 transformants produced avermectins. Although *S. avermitilis* is proficient for recombination, we believe that the production of avermectins was the result of trans complementation rather than recombination. On occasion we have seen results indicative of

recombination in which only 1/12 to 3/12 transformants produce avermectins. These putative recombinants were observed with only one or two members of a complementation class and only with a subclone derived from the integration vector.

FIG. 1 indicates the subclones which were used to successfully complement AGL$^-$ mutants. Table 1 identifies the mutants in each complementation class and presents the complementation results with key subclones. Twenty-one aglycone producing mutants, representing complementation Classes I, II, and III, were complemented after introduction of pVE650, but 5 Agl$^-$ mutants and two GMT- mutants were not. Class I mutants were complemented when they contained pVE650, or subclone pVE908 (2.4 *Eco*RI-*Bgl*II fragment). Class II mutants were complemented by pVE650 or subclone pVE807 (2.6 kb *Bgl*II fragment), but not by pVE908. Class III mutants were not complemented by pVE807 or pVE908. Although we can not exclude the occurrence of intragenic complementation, it is likely that each complementation class represents at least one gene for avermectin glycosylation. We have designated three genes to represent the loci defective within the mutants of complementation Classes I, II, and III, *avrB*, *avrD*, and *avrC*, respectively.

Isolation of cosmid clones which overlap pVE650 sequences

Since two avermectin genes were located to a 6.6 kb region at one end of pVE650, it was possible that the AGL$^-$ mutants which were not complemented by pVE650 might contain mutations in the DNA that maps adjacent to pVE650 in the *S. avermitilis* genome. To test this hypothesis the 1.1 kb *Bam*HI fragment from the end of the insert in pVE650 was used in a chromosome walk experiment to isolate overlapping clones from a *S. avermitilis* cosmid library. Colony hybridization to 2016 cosmid clones identified 5 cosmids. One cosmid, pVE855, contained all the DNA represented by pVE650 and additional DNA from each end. Collectively the cosmids represent 60 kb of *S. avermitilis* DNA. None of the cosmids overlapped sequences on pAT1. From one cosmid, pVE859, we identified a 15 kb *Bgl*II fragment which contained the 470 bp *Eco*RI to *Bam*HI fragment near the end of pVE650. Thus, this 15 kb fragment represents the chromosomal *Bgl*II fragment that is adjacent to the 140 bp *Bgl*II fragment of pVE650 and extends 13 kb beyond the DNA contained on pVE650. This fragment was cloned into pIJ922 to yield pVE941. pVE941 contains all the *S. avermitilis* DNA on pVE807 and, as expected, complements Class II aglycone producers. pVE941 also complemented all 5 AGL$^-$

mutants not complemented by pVE650 and two GMT$^-$ strains. Thus, the genes for glycosylation of avermectin are clustered since all the mutants defective in synthesis or addition of oleandrose to avermectin aglycone are complemented by pVE650 and/or pVE941.

Localization of additional genes for glycosylation of avermectin aglycone

Additional subclones were prepared from pVE855 and used in complementation tests. pVE1111 (4.1 *Eco*RI fragment of pVE650 plus the 1.8 kb *Eco*RI fragment of pVE941) complemented Class I, II and Class III mutants. Thus the mutants in Class III are be defective in a gene, designated *avr*C, located between *avr*B and *avr*D. MA6057 and MA6622 were complemented by only pVE941 and pVE1115 and are designated class IV. pVE1019, which contained the 3.5 kb *Bam*HI fragment from pVE941, complemented the defects in the two GMT mutants and AGL$^-$ strain MA6590. This later mutant was designated Class V. Two mutants complemented by pVE941 and pVE1018, but not by pVE650 or pVE1019, were designated Class VI. Table 1 summarizes the complementation results which have defined the 7 classes of mutants involved in glycosylation of the avermectin aglycone.

Subcloning a region which complements all AGL$^-$ mutants

An 12 kb *Pst*I fragment, which overlaps both pVE650 and pVE941, was subcloned onto pVE1043 to yield pVE1115. Mutants from all the complementation classes were complemented by pVE1115. Thus, it appears that all the genes for glycosylation of avermectin have been cloned on pVE1115. The 7 complementation classes define the minimum number of genes involved in avermectin glycosylation. Each complementation class may represent more than one gene. FIG. 1 shows the location of the 7 identified genes involved in glycosylation of avermectin.

Only AGL$^-$ mutants are complemented by pVE650

We tested pVE650 for the presence of other genes by complementation analysis. pVE650 was introduced into mutants representing each phenotypic class of *S. avermitilis* defective or altered in avermectin biosynthesis. No complementation was observed in MA6238 (C-22, C-23 dehydrase [DH$^-$]), MA5218 (C-6, C-8' furan ring formation [FUR$^-$]), MA6316 (C-3', C-3" O-methyltransferase or glycosyl O-

methyltransferase [GMT$^-$]), MA6262, (nonproducer of avermectin [NPA$^-$]), or MA6233 (OMT$^-$).

The complementation results with pVE1115 clearly show that the genes for glycosylation of avermectin are tightly clustered. pVE1115, which contains a 12 kb *Pst*I fragment from *S. avermitilis* , complemented all 26 mutants which fail to glycosylate avermectin and 2 mutants which fail to methylate hydroxyls at the C-3', C-3" positions. This suggests that pVE1115 may contain all the genes for synthesis and for attachment of oleandrose to avermectin aglycone. However, it is possible our collection of mutants does not include defects in all the genes involved in avermectin glycosylation. If this is so, then pVE1115 may not contain all the glycosylation genes.

Sequence of the glycosylation region.

BamHI, EcoRI, and PstI-BamHI fragments from pVE1101 were subcloned and sequenced on both strands using a primer walking strategy. DNA was sequenced manually using Sequenase (US Biochemicals) and an ABI 373A automated sequencer (Perkin Elmer) according to the manufacture's recommendations. The resulting 9994 nt sequence is shown as SEQ ID NO:1. And was analyzed by the GCG software suite (Genetics Computer Group). 9 complete ORF were identified and the genes involved in glycosylation designated AvrB through AvrI as shown on FIG. 1. In this region there are two sets of overlapping genes. The AvrB and AvrC genes are convergently transcribed and their coding regions overlap for 95 nt. The AvrD and AvrC genes are co-transcribed but encode proteins in different reading frames and overlap for 16 nts. A comparison of the open reading frames in the sequence to the clones used in complementation analysis results in the identification of 8 genes essential for avermectin glycosylation.

TFASTA comparison of the ORFs to Genbank resulted in highly significant similarities to several known genes. ORF1 showed similarity to keto-reductases. ORF2 showed greater than 30% identity to glycosyl-transferases. ORF3a was greater than 60% identical to TDP-glucose-4,6-dehydratases, ORF3b was greater than 60% identical to several TDP-glucose synthases, and ORF4 showed weak homology to keto reductases. ORF5 had greater than 50% identity to hexose 3,5 epimerases. ORF7 was identified as a glycosyl methyltransferase since that ORF could complement the GMT- mutants.

Macrolides contain many unusual sugars (Omura, S. Macrolide Antibiotics, Academic Press, 1984). A biochemical study of the mutants and cloned genes will help elucidate the biochemical pathway for synthesis of oleandrose. The cloned genes for synthesis and addition of oleandrose to avermectin aglycone can be useful in intergenic complementation studies to identify genes involved in glycosylation of other macrolides. Alternatively, the cloned DNA can be useful as a probe to identify genes involved in the synthesis and/or addition of other sugar moieties to other macrolides. For example, the *act*I gene of *S. coelicolor*, which is required for synthesis of actinorhodin, has been useful as a probe to identify putative polyketide synthetases from other species (Bergh and Uhlen, 1992).

The genes for glycosylation of the avermectin aglycone can be useful in the production of novel antibiotics. Since avermectins are much more potent antiparasitic agents than avermectin aglycones (Campbell, W. Ivermectin and Abamectin, Springer-Verlag, 1989) or the non-glycosylated, but similar milbemycins (Omura, S. Macrolide Antibiotics, Academic Press, 1984), it is evident that the oleandrose disaccharide moiety enhances the potency of avermectin. The genes described herein for synthesis and attachment of oleandrose to avermectin aglycone can be useful for the construction of hybrid antibiotics. For example, the introduction of a plasmid containing at least one gene of the present invention into strains that produce antibiotics with a hydroxyl group may result in hybrid glycosylated antibiotics. Potentially useful substrates for glycosylation are other macrolides (Omura, 1984).

REFERENCES

Baltz, R. H. and E. T. Seno. (1988) Genetics of *Streptomyces fradiae* and tylosin biosynthesis. Ann. Rev. Microbiol. **42**:547-574.

Bergh, S. and M. Uhlen. (1992) Analysis of a polyketide synthesis-encoding gene cluster of Streptomyces curacoi. Gene **117**:131-136

Burg, R. W., Miller, B. M., Baker, E. E., Birnbaum, J., Currie, S. A., Hartman, R., Kong, Y.L., Monaghan, R. L., Olson, G., Putter, I., Tunac, J. B., Wallic, H., Stapley, E.O., Oiwa, R. & Omura, S. (1979). Avermectins, new family of potent anthelmintic agents: producing organism and fermentation. Antimicrobial Agents and Chemotherapy **15**:361-367.

Butler, M.J., Friend, E.J., Hunter, I.S., Kaczmarek, F.S., Sugden, D.A., and Warren, M. (1989) Molecular cloning of resistance genes and architecture of a linked gene cluster involved in biosynthesis of oxytetracycline by *Streptomyces rimosus*. Mol. Gen. Genet. **215**:231-238.

Carter, G.T., J.A. Nietsche, M.R. Hertz, D. R. Williams, M.M. Siegel, G.O. Morton, J.C. James, and D. B. Borders. 1988. LL-F28249 antibiotic complex: A new family of antiparasitic macrocyclic lactones isolation, characterization, and structures of LL-F28249. J. Antibiotics **41**:519-529.

Chater, K.F., and D.A. Hopwood. 1984. *Streptomyces* genetics. In Biology of the Actinomycetes, pp229-286. Edited by M. Goodfellow, M. Mordarski, & S.T. Williams. London: Academic Press.

Epp, J. K., and B.E. Schoner. 1988. Molecular cloning and expression of carbomycin biosynthetic and resistance genes from *Streptomyces thermotolerans*. Abstr. Seventh Intern. Symp. Biol. Actinomycetes. S1b-2

Fisher, M. H., and H. Mrozik. 1984. The avermectin family of macrolide-like antibiotics. In Macrolide Antibiotics, pp. 553-604. Edited by S. Omura. Orlando, Fla.: Academic Press.

Goegelman, R.T., V. P. Gullo and L. Kaplan. 1983. Novel C-O76 compounds. U.S. Patent 4,378,353.

Hopwood, D.A. 1967. Genetic analysis and genome structure in *Streptomyces coelicolor*. Bacteriol. Reviews **31**:373-403.

Hopwood, D.A., M.J. Bibb, K.F. Chater, T. Kieser, C.J. Bruton, H.M. Kieser, D.J. Lydiate, C.P. Smith, J.M. Ward, and H. Schrempf. 1985. Genetic manipulations of *Streptomyces:* a laboratory manual. John Innes Foundation, Norwich, UK

Hopwood, D. A., T. Kieser, H. M. Wright, and M. J. Bibb. 1983. Plasmids, recombination, and chromosome mapping in *Streptomyces lividans* 66. J. Gen. Microbiol.**129**:2257-2269.

Ikeda, H., H. Kotaki, and S. Omura. 1987. Genetic studies of avermectin biosynthesis in *Streptomyces avermitilis*. J. Bacteriology **169**:5615-5621.

Ikeda, H., L. R.Wang, T. Ohta, J. Inokoshi, and S. Omura (1998) Cloning of the gene encoding avermectin B 5-O-methyltransferase in avermectin-producing *Streptomyces avermitilis*. Gene 12: 175-180.

Ikeda-H; Takada-Y; Pang-CH; Matsuzaki-K; Tanaka-H; Omura-S (1995) Direct production of 5-oxo derivatives of avermectins by a recombinant strain of Streptomyces avermitilis. J-Antibiot-Tokyo **48**(1): 95-7.

MacNeil, D.J. (1986). A flexible boiling procedure for isolating plasmid DNA from gram-positive microorganisms. Journal of Microbiological Methods **5**:115-124.

MacNeil, D.J. (1987). Introduction of plasmid DNA into *Streptomyces lividans* by electroporation. FEMS Microbiology Letters **42**:239-244

MacNeil. D.J. (1988) Characterization of a unique methyl-specific restriction system in *Streptomyces avermitilis*. J. Bacteriol. **170**:5607-5612

MacNeil, T & Gibbons, P.H. (1986). Characterization of the *Streptomyces* plasmid pVE1. Plasmid **16**:182-194.

MacNeil, D.J. and L.M. Klapko. (1987). Transformation of plasmid DNA into *Streptomyces avermitilis*. J. Industrial Microbiol. **2**:209-218.

MacNeil, D.J., Gewain, K.M., Ruby, C.L. Dezeny, G., Gibbons, P.H. and T. MacNeil (1992) Analysis of the *Streptomyces avermitilis* avermectin genes using a novel integration vector. Gene 111: 61-68

MacNeil, T., Gewain, K.M., and MacNeil, D.J. (1993) Deletion analysis of the avermectin biosynthetic genes of *Streptomyces avermitilis* by Gene Cluster Displacement. J. Bacteriol. 175: 2552-2563.

Malpartida, F., S.E. Hallam, H.M. Kieser, H. Motamedi, C.R. Hutchinson, M.J. Butler, D.A. Sugden, M. Warren, C. McKillop, C.R. Bailey, G.O. Humphreys and D. A. Hopwood. 1987. Homology between *Streptomyces* genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes. Nature (London) **325**:818-820.

Malpartida, F. and D. A. Hopwood. 1984. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. Nature (London) **309**:462-464.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Miller, J. H. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Motamedi, H. and C. R. Hutchinson. 1987. Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumer antibiotic of *Streptomyces glaucescens*. Proc. Natl. Acad. Sci. USA **84**:4445-4449

Murakami, T., H. Anzai, S. Imai, A. Satoh, K. Nagaoka, & C.J. Thompson. 1986. The Bialaphos biosynthetic genes of *Streptomyces hygroscopicus*: Molecular cloning and characterization of the gene cluster. Mol. Gen. Genet. **205**:42-50

Omura, S., *ed.* 1984. Macrolide Antibiotics: Chemistry, Biology, and Practice. Academic Press, Inc. Orlando, FL.

Putter, I., MacConnell, J.G., Preiser, F.A., Haidri, A.A., Ristich, S.S. & Dybas, R.A. 1981. Avermectins: novel insecticides, acaricides, and nematicides from a soil microorganism. Experientia **37**:963-964.

Ruby, C.L., M.D. Shulman, D.L. Zink & S.L. Streicher, 1986. Isolation and characterization of *Streptomyces avermitilis* mutants defective in methylation of the avermectins. In Biological., Biochemical and Biomedical Aspects of Actinomycetes. , *eds.* Szabo, G, S. Biro, & M. Goodfellow. Akademiai Kiado, Budapest. p279-280.

Schulman, M.D., D. Valentino, O. D. Hensens, D. Zink, M. Nallin, L. Kaplan, D. A. Ostlind. 1985. Demethylavermectins: biosynthesis, isolation and characterization. J. Antibiotics **38:**1494-1498.

Schulman, M.D and C. L. Ruby. 1987. "Streptomyces avermitilis" mutants defective in methylation of avermectins. Antimicr. Agents and Chemother. **31**:744-747.

Streicher, S. L., C. L. Ruby, P. S. Paress, J. B. Sweasy, S. J. Danis, D. J. MacNeil, K. Gewain, T. MacNeil, F. Foor, N. Morin, G. Cimis, R. Rubin, R. Goldberg, M. Nallin, M. D. Schulman, and P. Gibbons. 1989. Cloning genes for avermectin biosynthesis in *Streptomyces avermitilis.* In Genetics and Molecular Biology of Industrial Microorganisms, *eds.* C. L. Hershberger, S. W. Queener, and G. Hegeman. American Society for Microbiology, Washington, DC

Takiguchi, Y, H. Mishima, M. Okuda, and M. Terao. 1980. Milbemycins, a new family of macrolide antibiotics: fermentation, isolation, and physico-chemical properties. J. Antibiotics **33**:1120-1127.

Thompson, C. J., T. Keiser, J.M. Ward, and D.A. Hopwood. 1982. Physical analysis of antibiotic-resistance genes from *Streptomyces* and their use in vector construction. Gene **20**:51-62.

Stanzak, R., P. Matsushima, R. H. Baltz, and R. N. Rao. 1986. Cloning and expression in *Streptomyces lividans* of clustered erythromycin biosynthesis genes from *Streptomyces erythreus*. Bio/Technology **4**:229-232.

Table 1. Complementation of *S. avermitilis* aglycone producing mutants

Class	Mutants	pVE650	pVE908	pVE807	pVE941	pVE1019	pVE1018	pVE1115
I	GG900. MA6595. MA6586. MA6593, MA6056.MA6624	+	+	-	-	-	-	+
II	MA6582, GG898. MA6579. MA6581, MA6589, MA6591. MA5872	+	-	+	+	-	-	+
III	MA6278, MA6580. MA6583, MA6584, MA6585. MA6587. MA6588. MA6060	+	-	-	-	-	-	+
IV	MA6057. MA6622	-	-	-	+	-	-	+
V	MA6590	-	-	-	+	+	+	+
VI	MA6592. MA6594	-	-	-	+	-	+	+
GMT	MA6316. MA6323	-	-	-	+	+	-	+

Plasmids used are shown in FIG. 1. PVE650 has been described (MacNeil et al., 1992), pVE1115 contains the 11 kb PstI fragment which complements all avermectin aglycone producing mutants.

In most cases, at least 6 transformants of each plasmid into each mutant were tested for avermectin production by Microferm and TLC analysis.

WHAT IS CLAIMED:

1. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide encoding a polypeptide having an amino acid sequence selected from the group consisting of the eight amino acid sequences encoded by the polynucleotide sequence SEQ ID NO:1.

(b) a polynucleotide which is complementary to a polynucleotide of (a),

(c) a polynucleotide representing a polymorphic form of (a), and

(d) a polynucleotide comprising at least 20 nucleotides of the polynucleotide of (a), (b) or (c), said 20 nucleotides being highly specific for polynucleotide of (a).

2. The polynucleotide of claim 1 wherein the polynucleotide comprises nucleotides selected from the group consisting of natural, non-natural and modified nucleotides.

3. The polynucleotide of claim 1 wherein the internucleotide linkages are selected from the group consisting of natural and non-natural linkages.

4. The polynucleotide of claim 1 that includes the entire nucleotide sequence of SEQ ID NO:1.

5. The polynucleotide of claim 1 that includes at least a nucleotide sequence of the one of the open reading frames of SEQ ID NO:1.

6. The polynucleotide of claim 5 having a sequence of *Streptomycete* genomic DNA.

7. The polynucleotide of claim 5 having a sequence of an RNA.

8. An expression vector comprising a polynucleotide of claim 1.

9. A host cell comprising the expression vector of claim 8.

10. A process for expressing a protein encoded by a nucleic acid having the sequence of SEQ ID NO: 1 in a recombinant host cell, comprising:

(a) introducing an expression vector of claim 9 into a suitable host cell; and,

(b) culturing the host cells of step (a) under conditions which allow expression of said protein from said expression vector.

11. A substantially purified polypeptide having an amino acid sequence selected from the group consisting of

(a) a polypeptide having an amino acid sequence of encoded for by a nucleic acid having the sequence of SEQ ID NO:1, and

(b) a polypeptide representing a polymorphic form of (a).

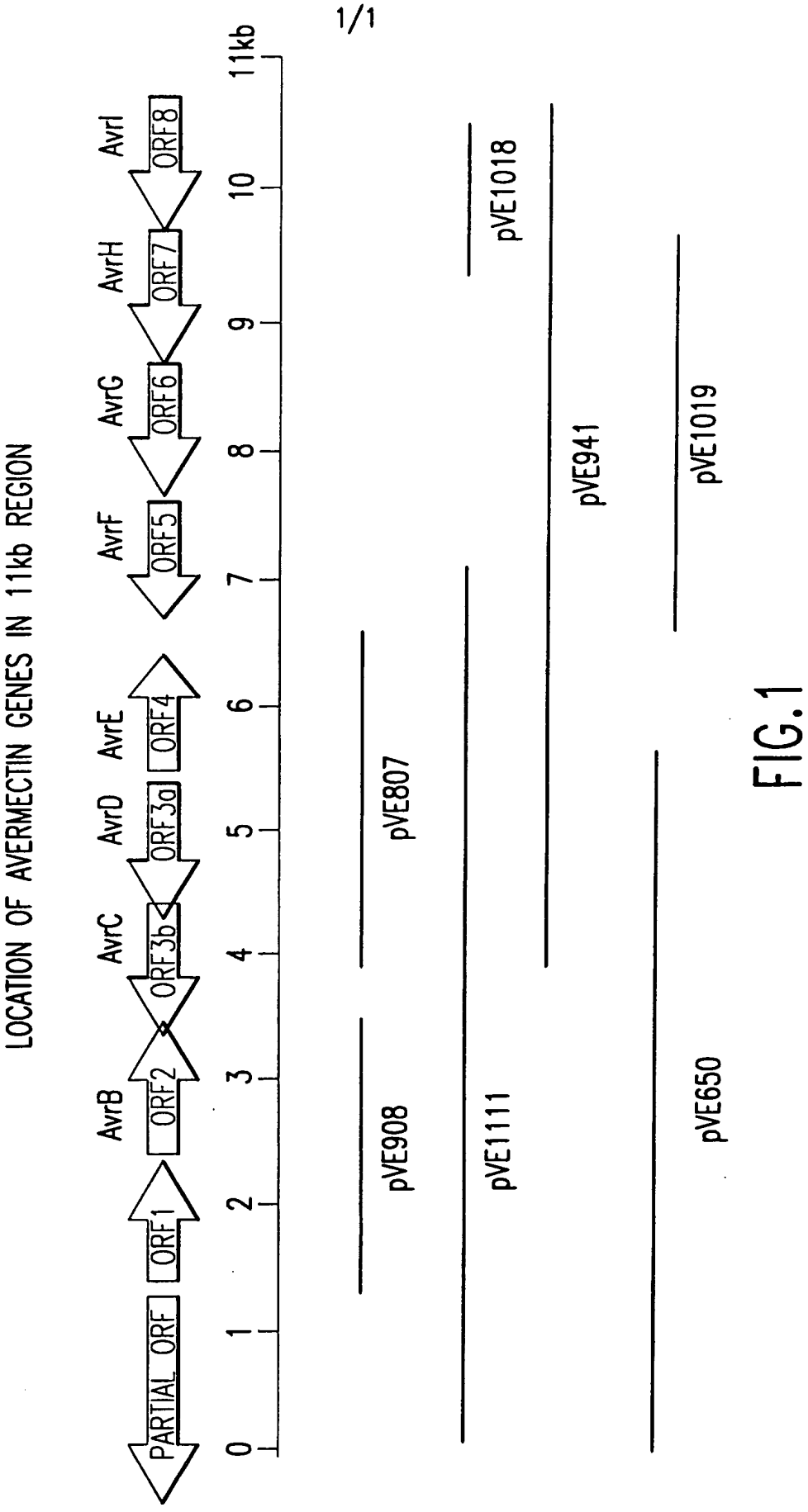
LOCATION OF AVERMECTIN GENES IN 11kb REGION
0
1
2
3
4
5
6
7
8
9
10
11kb
PARTIAL ORF
ORF1
AvrB
ORF2
AvrC
ORF3b
AvrD
ORF3a
AvrE
ORF4
AvrF
ORF5
AvrG
ORF6
AvrH
ORF7
AvrI
ORF8
pVE908
pVE807
pVE1111
pVE941
pVE1018
pVE650
pVE1019

FIG. 1

SEQUENCE LISTING

<110> Merck & Co., Inc.

<120> CLONING OF THE STREPTOMYCES AVERMITILIS
GENES FOR GLYCOSYLATION OF AVERMECTIN AGLYCONES

<130> 20506 PCT

<140> 60/146,699
<141> 1999-07-30

<160> 10

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 9994
<212> DNA
<213> Streptomyces avermitilis

<400> 1
ggatccatcg ccaacgcctc acgcggactg atcccgaaaa accccgcatc gaactccgcc 60
gcaccctcca ggaaaccgcc ccggcgcgta tacgacgaac ccgcccgccc cggctccgga 120
tcatagaaag cctccacgtc ccaaccccgg tcgaccggaa actcccccac cgcatcccga 180
cccgacgcaa tcaactccca gaaatcctcc gccgactcca cacccccccgg aaaacggcac 240
gccatcccca caattgcaat cggctcctgc tcgcccgatt caatctgctg aagtcgacgc 300
cgcacattga ggagatcggc agtaacgcgc ttgagatagt cgcggagctt ttcctcgtta 360
gccatggacc ggtctcctcg acaagagaaa tcggaaatta aaaaacacgc atgggactct 420
cacaggctag agcgacgaga gcagcacaaa tacccctaga taccccagac ccctgatgct 480
cgatgaatgc cgctatagct agggggtatg gcgccagaca tgaattcaca gcgtttcggc 540
ggccggctgg cgcttgtcac aggtgcaggc ggtggcatcg ggcgggcgac ctgcgctctc 600
ggatcggccg gggcgcgagt ggtgtgcgtg gaccgggacg gccgcggcgc cggggtgacg 660
ccgacctggc cggagcgggg cgcgcgggcg gcctggcccg aggtggccga cgtgtccgac 720
ggagcggcga tggagcggtt gcccgagcgc gtcgccgaga cgtacggggt cgtggacctg 780
ctggtgaaca acgccggcat cggcatggcg gggcgttttc tcgacacgtc cgtcgaggac 840
tggcagcgca ccctgggcgt caacctctgg ggtgtcattc atggttgccg cctcatcggc 900
cggcagatgg cggagcgcgg gcagggcggg cacatcgtga cggtggcgtc ggcggcggcg 960
ttccagccga cgcgggcggt ccccgcgtat gccaccagca aggcggcggt gctgatgctg 1020
agcgagtgcc tgcgcgcgga gttcgcggag ttcggggtcg gagtgagcgt ggtgtgcccg 1080
ggcttcgtcc gtacgtcgtt cgcgtcggcg atgcatttcg ccggtgtgcc ccggctggag 1140
caggagcggc tgcgggcgct gttcgccggt cgcggatgca gcgcggagaa ggtggccgcg 1200
gcggtactgc ggtcggtggc gcgcgactcg gccgtggtga ccgtgacggc ggaagcgcgg 1260
ctgtcacggc tgatgagccg cttcacgcca cgcctgcgcg ccgcggtggc gcggatggat 1320
ccccccttcgt agggctggcg gggatcccct ccttgccttc gaacatcttc cgacgatggg 1380
cagtgagaga tgtcagatca ttttctcttc atgagtgcgc cgttctgggg gcatgtgttc 1440
cccagtctcg ccgtggcgga ggagctcgtg caccggggcc accacgtcac ctttgtgacg 1500
ggcgcggaaa tggccgatgc ggtgcgttcc gtgggcgctg atttcctgcg gtacgagtcc 1560
gccttcgagg gtgtcgacat gtaccggctg atgaccgagg ccgagccgaa cgccatcccc 1620
atgacgctgt acgacgaggg catgtccatg ttgcgttcgg tggaggagca cgtcggcaag 1680
gacgttccgg acctggtggc ctacgacatc gccacctccc tcaacgtggg tcgtgtcctc 1740
gccgcctcct ggagcaggcc ggccatgacg gtcattcccc tgttcgcgtc caacgggcgc 1800
ttctccacga tgcagtcggt attggatccg gattccgctc aggtcagtgc gccgccgccg 1860
cgcttctcgg agcagatgga gttgttcggc ctcggggcgc tggtgccgcg cctcgcggag 1920
ctgctcgttt cccggggtat cacggaaccg gtcgacgatt tcctttccgg accggaggac 1980
ttcaacctgg tgtgtctgcc gcgcgccttc cagtacgcgg gcgacacctt cgacgagcgg 2040
ttcgccttcg tcggaccatg tctgggtaag cgcaggggtc tgggcgagtg gacaccaccg 2100
ggcagcgggc atccagtggt gctcatctcc ctcgggaccg tgttcaaccg gcagctgtcc 2160
ttcttccgca cgttcgtccg ggcgttcacc gacgtccccg tgcacgtcgt gatctcgctc 2220
ggcaaggggg tcgaccccga tgtgctgcgg ccgctgccgc cgaatgtcga ggtgcaccgg 2280

```
tgggtgccgc accatgcggt gctggagcat gccagggctc tggtcacgca cggcggtacc  2340
ggcagtgtga tggaggcact gcacgcaggg tgcccggtgc tcgtcatgcc cttgtcgcgg  2400
gacgcgcagg tgaccggccg gcggatcgcc gagctggggc tgggtcgtat ggtgcagccg  2460
gaggaggtca cggcgacgac gctgcgccgg cacgtgctgg acatcatctc cgatgacgcg  2520
atcacccgac aggtcaggca gatgcagcgg gccacggtcg aggcgggcgg cgccctgcgg  2580
gcagcggacg agaccgagcg gtttctgcgc cggacgcgcc gtcactgacc ggcagctcgg  2640
gccgggcggt gagtggctcc cacagggttc ggttctccac gtaccactga acggtctgtg  2700
ccagcccctc ctcgaagggc acgcggggcg cgtaaccgag ctcggcggag atcttgctga  2760
tgtccagcga gtagcgccgg tcgtgcccct tgcggtcggt cacgggttcg accatcgacc  2820
agtccacgcc gagcaggtcc aggagccggg cggtgagctc acggttggac agctccgtcc  2880
cgcctccgat gtggtagatc tcgccgggcc tgtcgcgttc ggcgaccagg gcgatgccac  2940
ggcagtggtc gtccacgtgc agccagtcgc ggacgttttc gccgtcgccg tacaagggca  3000
ccttcgtgcc gttcagcaga tgggtgacga accgcgggat gagtttctcc gggaactggt  3060
gggggccgta gttgttcgag catcgggtga tgatcactgg taggccgtgc gtgcggtgga  3120
aggaccgggc gagcaggtcg gaggacgcct tggacgcgga gtagggcgag ttcggctcca  3180
gcggggcgtc ctcggtccac gagccggagt cgatggagcc gtagacctcg tccgtcgaga  3240
tgtacacgaa gcggtccacg gcggcgtcgg tggcggcgcg gagcagggtg tgagtgccga  3300
ggacattggt gcgtacgaac tcggcggcgt cggccacgga ccggtccacg tgtgactccg  3360
ccgcgaagtg gaccaccatg tcggagccgt ccatcaggtc cgcgaccaag ggcccgtcgc  3420
agatgtcgcc gtgcacgaag atcagggatg ggcttcccag gaccggtgcg aggttctcca  3480
ggcgacccgc gtaggtcagc ttgtcgagca ccacgacctc ggcaccggtg aacgccggat  3540
acgcgcccgt cagcaaccgc cgtacgaaat gggaaccgat gaaaccggcg ccgcccgtca  3600
cgagtaggcg catcccgggc tcctcaccgc ggcttccgcc gcaatactca tcagatactc  3660
gccgtagccg gagccggcca gttcgacccc gcgcagatag cagtcgtccg cgtcgatcag  3720
acccatccgg aaggcgatct cctcgagaca ggcgatccgt actccctggc gcttctccag  3780
gacctgcaca tactgcccgg cgtgcatcag cgagtcgtgc gtccccgcat cgagccaggt  3840
gaagccccgg cccaggtcca ccagccgggc ccgcccctcg gcgaggtagg ccctgttgac  3900
gtcggtgatc tccagctcgc cgcgggccga cgagcggatg ccccgggcca cctcgatcac  3960
gtcgttgtcg tacaggtaca ggcctgtgat cgccaggttg gaccgggggg cggtgggttt  4020
ctcctcgacg gacagcagct ttccggaggc gtcgacctct ccgactccgt accgttcggg  4080
atccgtcacc gcgtatccga acaacacaca gccgtcgaca tcgcgggtgt ggctgcgcag  4140
caggtgcgaa aagcccatgc catggaagat gttgtcccca aggacaaggg acacctgatc  4200
ctgaccgatg aaatcggcgc cgatgaggaa tgcctcggcg attcctcccg gtcgctgctg  4260
cgcggcgtag tcgatgttca gcccgaggcg gcttccgtct ccgagcagtc tccggaattg  4320
ttcgagatga tcgggtgagg aaatcaccag gatgtctttt atgccgccga gcatcaacac  4380
ggagagcggg tagtagatca tgggtttgtc gtagacaggg agcagctgct tggaaagggc  4440
acgggtcaac gggtaaagcc gagagccggt tccccccgcg agcacgattc ccttcatgtc  4500
ggactccccg cagtcgacgt tatatatctc gtgccgtctg cccgacggta ccaagtggcg  4560
gaaaacgcac caggaattcg agcgccgcta gggggaaggg ctcaagaaga taggggccac  4620
cagatggggc ggttttcggt gtgcccgccc cggccgaccg gaatactgaa gagcatgctg  4680
acgactggga tgtgcgaccg accgctggtc gtcgtactcg gagcctccgg ctatatcggg  4740
tcggccgtcg cggcggaact cgcccggtgg ccggtcctgt tgcggctggt ggcccggcga  4800
ccgggcgtcg ttccgccggg cggcgccgcg gagaccgaga cgcgtacggc cgacctgacg  4860
gcggcgagcg aggtcgccct cgccgtgacg gacgccgacg tggtgatcca cctggtcgcg  4920
cgcctcaccc agggagcggc atggcgggcg gcggagagcg atccggtggc cgagcgggtg  4980
aacgtcgggg tgatgcacga cgtcgtcgcg gccctgcggt ccgggcgccg cgccgggccg  5040
cccccggtgg tggtgttcgc cgggtcggtc taccaggtgg gccgcccggg tcgggtcgac  5100
ggcagtgagc cggacgagcc cgtgacggcc tatgcccgtc agaaactcga cgccgaacgg  5160
acgttgaagt ccgccacggt cgagggtgtc ctgcggggga tctcgctgcg gctgcccacc  5220
gtctacggcg cggggccggg cccgcagggc aacggcgtcg tgcaggcgat ggtgctccgg  5280
gcgctcgccg acgaggccct caccgtgtgg aacggaagcg tggtggagcg tgacctggtg  5340
catgtggagg atgtcgcgca ggccttcgtg agctgcctgg cgcacgcgga tgcgctcgcc  5400
gggcggcact ggctgctcgg cagcggtcgt cctgtgaccg tcccgcacct cttcggtgcc  5460
atcgccgccg gcgtgtccgc ccgcaccggg cgccccgcgg tgcccgtgac cgcggtggac  5520
cctccggcga tggcgacggc ggcggacttc cacgggaccg tcgtcgactc ctcggcgttc  5580
cgcgcggtca ccgggtggcg gccgcggctg tcgcttcagg agggcctgga ccacatggtg  5640
gcggcttacg tgtagcgccg gggtggcggc cgggcccggg cggtgacggc ccggatccgg  5700
gtcggccgtc acagcttctc gtcgagggcg gggctcgcgc ggtactccgg caacatgccg  5760
cgtcgcaggg cctgctggag agtcggcgcg cgcgccggtc cgcgctcgga gaggatcggt  5820
gcccgcccga ggtggtggcc gaggggcagg gcgaggtccg gatcctcggg cgagagggcg  5880
tgttcgttct gcggaacgta gccgctcgac atcaggtaca ccatcgccgt gtcgtcttcc  5940
agcgccacga acgcgtgccc gaccccgatc ggcaggtaga cggaacggaa gcgctcctgg  6000
```

```
tcgaggagga ccgagtccca ctgcccgaaa gtcggtgagc cggtgcgcag gtcgacgacg  6060
aagtccaggg cccgtccccg ggcgcagtgg acgtacttgg cctggccggg tggtgtcgcg  6120
gtgaagtgca cgccgcggac gacgccgcgg cgcgagacgc tctggcaggt ctgcgcggtg  6180
ggaaaccggt gcccgacggc ctcgctgagg accggttcct ggtagggggt gacgaagagc  6240
ccgcgctcgt cggggaagac cgtcggggtg aattcgacgg cgccctcgac gacgagcctc  6300
cggaccgtga caccggcggc ggtggcccgg gcgcccgcgg gcggggcggg ccggtcggcg  6360
gagctccggc gaggccggcc aagggtcatc gctgcactct ctctgtcgtg cgggttgtca  6420
tacgggtagt cgtacgggcc ggttccggag tcacagctcg acggcgcggg tggtgagcag  6480
ggacagcagg gtgcgggcct gcacgttcac gtaacggccg taccgcagca gctgggtcag  6540
ctggcccggg gtgcaccagc ggtaccccgg gggcgggtcg ttcggcgcct ggctctcgtc  6600
ggcctcgacg aacaggtagc gcgcctgtgc gtgcagaaag cgaccgccct cctccgagtg  6660
gaccgccgcg tagcggatgc ggtcgggcgc ggcctccagc accaggtcga ggaagcgcgg  6720
cctggccggt cccgtgaggt gggcgtagtt gcgcggggtg tactggaccg tcgggccgag  6780
ttcgatcgtg tcgaggaagc cgccctcgac cctgccgtgg gcgagcaggt gcggtacgcc  6840
gccgatccgc cgggtcagga aggcggtgat gccgtggccg cacggttcga tcaggggctg  6900
ggtccaggcg gcgacctccc ggttggaggc ctcgacacgg accgcgacca cacggaagta  6960
ccggtccgcg tggtgggcga tggactccgc gcccgtggtc cagccgggga tgccggccag  7020
gggcacgcgg cgggcgtgca cggagtgccg ggagcgttcg gcggcgtacc aggagagcag  7080
ttcggcgtcg ctgtgcaggg ccgcgggctc gtcgaacggg gtgggaaggc aggcgaggac  7140
cgtgcgtgcg tccatgttca ccaggttgtc ccggtgcatc agttcgccga tctgccccag  7200
tgtcagccag cggaagtcgt cgtccagtgg tacgtcctcg tcggtctcca ccacgatgtt  7260
gcggttgaac ttccggtgga accaggctcc gtgctcggac tggaggacgt cgaccaccac  7320
ggtggcgcgc cggggctgtg tgaagtactc gaggtacttc acggcggcgc ccccgtggac  7380
cttggtgtag ttgctgcgcg tggcctgcac ggtgggcgac agctggacca ggttgatgtt  7440
gccgggctcc atcttggcct gcatcaggaa gtgcaggacc ccgtcgaact tcttggcgag  7500
gatgccgagg atgccgatct cgggctggtg gatgatgggc tgctgccatt ccgggaaggg  7560
ctgttcaccg cctcggacgt gcagtccctc cacggagaag aaccggccgc tgcggtgggc  7620
cagattgccg gttccggggt gaaacgacca ggcgtccatc ccgtggaagg ggatgcgctc  7680
gacccggaac cggtgggccc cggaccgccg cgtccaccag ccggtgaacg cgtcgaggga  7740
cgtccgggcg ccggtgtcgc ccacggcggc ggagcgggcg aggcacgcgg gcagggcggc  7800
gtcgtgccgc gcggtgagcg gtgctgggct cggtgtggtc ggcatcggct cgtacgctca  7860
tgcaccccac gtcatgtaga tcaccggtgg ctcgcggccg ggcagttggc gcagtggggc  7920
gtggtcgagg ccgaacgcct cgctcagcgc cctggtctcc cccggccatt tggggtgggt  7980
gagttcgtcg aaggcgagga tgctgccccc ggtcaggtgc ggtgtgatga cgtccagcag  8040
ttcgcgcgtg gggcggtaga ggtccaggtc gaagtaggcc agcgcgatga cggtgtgcgg  8100
gtgttccgcc aggtattggg gcaccgtttc gcgtacgtcg ccctggacca cgaaggaacg  8160
ctgggtgtgg ccgtagggtt cgttcgcctc gtgcgccgcg agcacctgcc gcaggtgctc  8220
cacttcgccg tccggcacgg cgaaccgccc agggaccgcg ctggtgctga cctcgtccgc  8280
ctcgtcgatg tcggggaagc cggtgaacgt gtcgaagccg atgacgcggc gcagcgagtt  8340
gtacggctca tagatgctgc gcagcgcggt cagcgtggcg aggtgccgtc cgtgcagaac  8400
gccgaactcc atgatgacgc cggggacttc cggcagcatg cggtacagcg cgtccatgga  8460
gagcaggtcg gcgagctggt tgcgccgcat gtagacggac aggttgtcga tcaggtactt  8520
cggcgggatc gggctgtcga cgaggagctt ggtcagctgc tcgcgggcag cgcgttcctg  8580
ctcggactcg tgcggcacga tccggggatc ggtgaactcc cgctcggtca tggaggcctt  8640
tcctttcatg ggtcggtacc gggcgcgccg gacgtgccgg tcgtaccggg cgtgccggcg  8700
ggcacgacgc tgtcgggtca ggacagccag gcgtcggggg cggatccgcc gcggccgacc  8760
gggggggaaca gctcctccag gcgggccagg acgggctcgg gcagcggggt gcgcagggcg  8820
tgcagtgccc cgtccacgtg ctgttcggtg cgcggcccga tgaccagccc ggtcacgccg  8880
ggccgcgaca gcacccaggc catgccgaca tgggcggggt cgaggccgtg gtccgcgcac  8940
acgtcctcgt acgccgcgat ggtggtgcgg tggtgctcca gggcctcgac ggcccggccc  9000
tgtgccgact tgaccgcggt gttctcccgc gtcttgcgca ggacaccgcc gagcaggccg  9060
ccgtgcagtg gcgaccagac caggacgccg acaccgtagg cggacgcggc ggggatgact  9120
tccagctcgg cgtgtcgggt cacgaggttg tagacgcact gctcggaggc gaggcccagg  9180
gcgttgcgcc gccgggccgc ctcctgggcg gaagcgatgt cccagcccgc gaagttggag  9240
gagccgacgt agcgcacctt gccctgcgtg atgagcaggt ccatcgcctg ccacacctcg  9300
tcccagccgg cgcggcggtc gatgtggtgc agctggtaca ggtcgatcca gtcggtgcgc  9360
agtcggcgca gcgaggcgtc gcaggcggcc acgatattgc gtacggacag tccgtgatcg  9420
ttggggccgc tgcccatcgg atcgccgacc ttggtggcca gcaccacctg ctcacgccgg  9480
gcggggcggt ccgccagcca cctgccgatg acctcttcgg tgtacccctt gtggacgcgc  9540
cagccgtagg tgttggcggt gtcgaacagg gtgatgccct gagccagggc gtgatccatc  9600
agtcggcgcg cttcgggctc ctccacccgt ccgccgatgt tgaccgttcc gagcgccagt  9660
cggctgatcc tcagccgggt cctgcccagt tcggtgtgga ggggagcact gctgttgctg  9720
```

```
tcggactgga cgggtgcggg ctcggccgtc gtaggcatca tcgatcagtc gacactccct    9780
cgtgcgtgag cggcgggcgc tcgagcagga ccctgacctg aggcccagga ggctaccggc    9840
gatcatgcga tacaggcagc cgctcgatgg tgggacacgg gctgccgtcg ccgggcatag    9900
gggctgatgg gggttgtccg gtgcgggtcc ggctgacagc ctcgtggaca ccaagttgat    9960
ccagttgatc cactccgaaa ggcagaggct gcag                                9994
```

<210> 2
<211> 274
<212> PRT
<213> Streptomyces avermitilis

<400> 2

```
Met Ala Pro Asp Met Asn Ser Gln Arg Phe Gly Gly Arg Leu Ala Leu
 1               5                  10                  15
Val Thr Gly Ala Gly Gly Gly Ile Gly Arg Ala Thr Cys Ala Leu Gly
            20                  25                  30
Ser Ala Gly Ala Arg Val Val Cys Val Asp Arg Asp Gly Arg Gly Ala
        35                  40                  45
Gly Val Thr Pro Thr Trp Pro Glu Arg Gly Ala Arg Ala Ala Trp Pro
    50                  55                  60
Glu Val Ala Asp Val Ser Asp Gly Ala Ala Met Glu Arg Leu Pro Glu
65                  70                  75                  80
Arg Val Ala Glu Thr Tyr Gly Val Val Asp Leu Leu Val Asn Asn Ala
                85                  90                  95
Gly Ile Gly Met Ala Gly Arg Phe Leu Asp Thr Ser Val Glu Asp Trp
            100                 105                 110
Gln Arg Thr Leu Gly Val Asn Leu Trp Gly Val Ile His Gly Cys Arg
        115                 120                 125
Leu Ile Gly Arg Gln Met Ala Glu Arg Gly Gln Gly Gly His Ile Val
    130                 135                 140
Thr Val Ala Ser Ala Ala Ala Phe Gln Pro Thr Arg Ala Val Pro Ala
145                 150                 155                 160
Tyr Ala Thr Ser Lys Ala Ala Val Leu Met Leu Ser Glu Cys Leu Arg
                165                 170                 175
Ala Glu Phe Ala Glu Phe Gly Val Gly Val Ser Val Val Cys Pro Gly
            180                 185                 190
Phe Val Arg Thr Ser Phe Ala Ser Ala Met His Phe Ala Gly Val Pro
        195                 200                 205
Arg Leu Glu Gln Glu Arg Leu Arg Ala Leu Phe Ala Gly Arg Gly Cys
    210                 215                 220
Ser Ala Glu Lys Val Ala Ala Ala Val Leu Arg Ser Val Ala Arg Asp
225                 230                 235                 240
Ser Ala Val Val Thr Val Thr Ala Glu Ala Arg Leu Ser Arg Leu Met
                245                 250                 255
Ser Arg Phe Thr Pro Arg Leu Arg Ala Ala Val Ala Arg Met Asp Pro
            260                 265                 270
Pro Ser
```

<210> 3
<211> 412
<212> PRT
<213> Streptomyces avermitilis

<400> 3

```
Met Ser Asp His Phe Leu Phe Met Ser Ala Pro Phe Trp Gly His Val
 1               5                  10                  15
Phe Pro Ser Leu Ala Val Ala Glu Glu Leu Val His Arg Gly His His
            20                  25                  30
Val Thr Phe Val Thr Gly Ala Glu Met Ala Asp Ala Val Arg Ser Val
```

```
        35                  40                  45
Gly Ala Asp Phe Leu Arg Tyr Glu Ser Ala Phe Glu Gly Val Asp Met
    50                  55                  60
Tyr Arg Leu Met Thr Glu Ala Glu Pro Asn Ala Ile Pro Met Thr Leu
65                  70                  75                  80
Tyr Asp Glu Gly Met Ser Met Leu Arg Ser Val Glu Glu His Val Gly
                85                  90                  95
Lys Asp Val Pro Asp Leu Val Ala Tyr Asp Ile Ala Thr Ser Leu Asn
            100                 105                 110
Val Gly Arg Val Leu Ala Ala Ser Trp Ser Arg Pro Ala Met Thr Val
        115                 120                 125
Ile Pro Leu Phe Ala Ser Asn Gly Arg Phe Ser Thr Met Gln Ser Val
    130                 135                 140
Leu Asp Pro Asp Ser Ala Gln Val Ser Ala Pro Pro Pro Arg Phe Ser
145                 150                 155                 160
Glu Gln Met Glu Leu Phe Gly Leu Gly Ala Leu Val Pro Arg Leu Ala
                165                 170                 175
Glu Leu Leu Val Ser Arg Gly Ile Thr Glu Pro Val Asp Asp Phe Leu
            180                 185                 190
Ser Gly Pro Glu Asp Phe Asn Leu Val Cys Leu Pro Arg Ala Phe Gln
        195                 200                 205
Tyr Ala Gly Asp Thr Phe Asp Glu Arg Phe Ala Phe Val Gly Pro Cys
    210                 215                 220
Leu Gly Lys Arg Arg Gly Leu Gly Glu Trp Thr Pro Pro Gly Ser Gly
225                 230                 235                 240
His Pro Val Val Leu Ile Ser Leu Gly Thr Val Phe Asn Arg Gln Leu
                245                 250                 255
Ser Phe Phe Arg Thr Phe Val Arg Ala Phe Thr Asp Val Pro Val His
            260                 265                 270
Val Val Ile Ser Leu Gly Lys Gly Val Asp Pro Asp Val Leu Arg Pro
        275                 280                 285
Leu Pro Pro Asn Val Glu Val His Arg Trp Val Pro His His Ala Val
    290                 295                 300
Leu Glu His Ala Arg Ala Leu Val Thr His Gly Gly Thr Gly Ser Val
305                 310                 315                 320
Met Glu Ala Leu His Ala Gly Cys Pro Val Leu Val Met Pro Leu Ser
                325                 330                 335
Arg Asp Ala Gln Val Thr Gly Arg Arg Ile Ala Glu Leu Gly Leu Gly
            340                 345                 350
Arg Met Val Gln Pro Glu Glu Val Thr Ala Thr Thr Leu Arg Arg His
        355                 360                 365
Val Leu Asp Ile Ile Ser Asp Asp Ala Ile Thr Arg Gln Val Arg Gln
    370                 375                 380
Met Gln Arg Ala Thr Val Glu Ala Gly Gly Ala Leu Arg Ala Ala Asp
385                 390                 395                 400
Glu Thr Glu Arg Phe Leu Arg Arg Thr Arg Arg His
                405                 410
```

<210> 4
<211> 359
<212> PRT
<213> Streptomyces avermitilis

```
<400> 4
Met Arg Leu Leu Val Thr Gly Gly Ala Gly Phe Ile Gly Ser His Phe
 1                  5                  10                  15
Val Arg Arg Leu Leu Thr Gly Ala Tyr Pro Ala Phe Thr Gly Ala Glu
            20                  25                  30
Val Val Val Leu Asp Lys Leu Thr Tyr Ala Gly Arg Leu Glu Asn Leu
        35                  40                  45
Ala Pro Val Leu Gly Ser Pro Ser Leu Ile Phe Val His Gly Asp Ile
    50                  55                  60
```

```
Cys Asp Gly Pro Leu Val Ala Asp Leu Met Asp Gly Ser Asp Met Val
65                  70                  75                  80
Val His Phe Ala Ala Glu Ser His Val Asp Arg Ser Val Ala Asp Ala
                85                  90                  95
Ala Glu Phe Val Arg Thr Asn Val Leu Gly Thr His Thr Leu Leu Arg
            100                 105                 110
Ala Ala Thr Asp Ala Ala Val Asp Arg Phe Val Tyr Ile Ser Thr Asp
        115                 120                 125
Glu Val Tyr Gly Ser Ile Asp Ser Gly Ser Trp Thr Glu Asp Ala Pro
    130                 135                 140
Leu Glu Pro Asn Ser Pro Tyr Ser Ala Ser Lys Ala Ser Ser Asp Leu
145                 150                 155                 160
Leu Ala Arg Ser Phe His Arg Thr His Gly Leu Pro Val Ile Ile Thr
                165                 170                 175
Arg Cys Ser Asn Asn Tyr Gly Pro His Gln Phe Pro Glu Lys Leu Ile
            180                 185                 190
Pro Arg Phe Val Thr His Leu Leu Asn Gly Thr Lys Val Pro Leu Tyr
        195                 200                 205
Gly Asp Gly Glu Asn Val Arg Asp Trp Leu His Val Asp Asp His Cys
    210                 215                 220
Arg Gly Ile Ala Leu Val Ala Glu Arg Asp Arg Pro Gly Glu Ile Tyr
225                 230                 235                 240
His Ile Gly Gly Gly Thr Glu Leu Ser Asn Arg Glu Leu Thr Ala Arg
                245                 250                 255
Leu Leu Asp Leu Leu Gly Val Asp Trp Ser Met Val Glu Pro Val Thr
            260                 265                 270
Asp Arg Lys Gly His Asp Arg Arg Tyr Ser Leu Asp Ile Ser Lys Ile
        275                 280                 285
Ser Ala Glu Leu Gly Tyr Ala Pro Arg Val Pro Phe Glu Glu Gly Leu
    290                 295                 300
Ala Gln Thr Val Gln Trp Tyr Val Glu Asn Arg Thr Leu Trp Glu Pro
305                 310                 315                 320
Leu Thr Ala Arg Pro Glu Leu Pro Val Ser Asp Gly Ala Ser Gly Ala
                325                 330                 335
Glu Thr Ala Arg Ser Arg Pro Leu Pro Ala Gly Arg Arg Pro Pro Arg
            340                 345                 350
Pro Trp Pro Ala Ala Ser Ala
        355
```

<210> 5
<211> 299
<212> PRT
<213> Streptomyces avermitilis

```
<400> 5
Met Lys Gly Ile Val Leu Ala Gly Gly Thr Gly Ser Arg Leu Tyr Pro
 1                  5                  10                  15
Leu Thr Arg Ala Leu Ser Lys Gln Leu Leu Pro Val Tyr Asp Lys Pro
            20                  25                  30
Met Ile Tyr Tyr Pro Leu Ser Val Leu Met Leu Gly Gly Ile Lys Asp
        35                  40                  45
Ile Leu Val Ile Ser Ser Pro Asp His Leu Glu Gln Phe Arg Arg Leu
    50                  55                  60
Leu Gly Asp Gly Ser Arg Leu Gly Leu Asn Ile Asp Tyr Ala Ala Gln
65                  70                  75                  80
Gln Arg Pro Gly Gly Ile Ala Glu Ala Phe Leu Ile Gly Ala Asp Phe
                85                  90                  95
Ile Gly Gln Asp Gln Val Ser Leu Val Leu Gly Asp Asn Ile Phe His
            100                 105                 110
Gly Met Gly Phe Ser His Leu Leu Arg Ser His Thr Arg Asp Val Asp
        115                 120                 125
Gly Cys Val Leu Phe Gly Tyr Ala Val Thr Asp Pro Glu Arg Tyr Gly
```

```
    130                 135                 140
Val Gly Glu Val Asp Ala Ser Gly Lys Leu Leu Ser Val Glu Glu Lys
145                 150                 155                 160
Pro Thr Ala Pro Arg Ser Asn Leu Ala Ile Thr Gly Leu Tyr Leu Tyr
                165                 170                 175
Asp Asn Asp Val Ile Glu Val Ala Arg Gly Ile Arg Ser Ser Ala Arg
            180                 185                 190
Gly Glu Leu Glu Ile Thr Asp Val Asn Arg Ala Tyr Leu Ala Glu Gly
        195                 200                 205
Arg Ala Arg Leu Val Asp Leu Gly Arg Gly Phe Thr Trp Leu Asp Ala
    210                 215                 220
Gly Thr His Asp Ser Leu Met His Ala Gly Gln Tyr Val Gln Val Leu
225                 230                 235                 240
Glu Lys Arg Gln Gly Val Arg Ile Ala Cys Leu Glu Glu Ile Ala Phe
                245                 250                 255
Arg Met Gly Leu Ile Asp Ala Asp Asp Cys Tyr Leu Arg Gly Val Glu
            260                 265                 270
Leu Ala Gly Ser Gly Tyr Gly Glu Tyr Leu Met Ser Ile Ala Ala Glu
        275                 280                 285
Ala Ala Val Arg Ser Pro Gly Cys Ala Tyr Ser
    290                 295

<210> 6
<211> 343
<212> PRT
<213> Streptomyces avermitilis

<400> 6
Met Gly Arg Phe Ser Val Cys Pro Pro Arg Pro Thr Gly Ile Leu Lys
 1                5                  10                  15
Ser Met Leu Thr Thr Gly Met Cys Asp Arg Pro Leu Val Val Val Leu
            20                  25                  30
Gly Ala Ser Gly Tyr Ile Gly Ser Ala Val Ala Ala Glu Leu Ala Arg
        35                  40                  45
Trp Pro Val Leu Leu Arg Leu Val Ala Arg Arg Pro Gly Val Val Pro
    50                  55                  60
Pro Gly Gly Ala Ala Glu Thr Glu Thr Arg Thr Ala Asp Leu Thr Ala
65                  70                  75                  80
Ala Ser Glu Val Ala Leu Ala Val Thr Asp Ala Asp Val Val Ile His
                85                  90                  95
Leu Val Ala Arg Leu Thr Gln Gly Ala Ala Trp Arg Ala Ala Glu Ser
            100                 105                 110
Asp Pro Val Ala Glu Arg Val Asn Val Gly Val Met His Asp Val Val
        115                 120                 125
Ala Ala Leu Arg Ser Gly Arg Arg Ala Gly Pro Pro Pro Val Val Val
    130                 135                 140
Phe Ala Gly Ser Val Tyr Gln Val Gly Arg Pro Gly Arg Val Asp Gly
145                 150                 155                 160
Ser Glu Pro Asp Glu Pro Val Thr Ala Tyr Ala Arg Gln Lys Leu Asp
                165                 170                 175
Ala Glu Arg Thr Leu Lys Ser Ala Thr Val Glu Gly Val Leu Arg Gly
            180                 185                 190
Ile Ser Leu Arg Leu Pro Thr Val Tyr Gly Ala Gly Pro Gly Pro Gln
        195                 200                 205
Gly Asn Gly Val Val Gln Ala Met Val Leu Arg Ala Leu Ala Asp Glu
    210                 215                 220
Ala Leu Thr Val Trp Asn Gly Ser Val Val Glu Arg Asp Leu Val His
225                 230                 235                 240
Val Glu Asp Val Ala Gln Ala Phe Val Ser Cys Leu Ala His Ala Asp
                245                 250                 255
Ala Leu Ala Gly Arg His Trp Leu Leu Gly Ser Gly Arg Pro Val Thr
            260                 265                 270
```

```
Val Pro His Leu Phe Gly Ala Ile Ala Ala Gly Val Ser Ala Arg Thr
        275                 280                 285
Gly Arg Pro Ala Val Pro Val Thr Ala Val Asp Pro Pro Ala Met Ala
    290                 295                 300
Thr Ala Ala Asp Phe His Gly Thr Val Val Asp Ser Ser Ala Phe Arg
305                 310                 315                 320
Ala Val Thr Gly Trp Arg Pro Arg Leu Ser Leu Gln Glu Gly Leu Asp
                325                 330                 335
His Met Val Ala Ala Tyr Val
            340
```

<210> 7
<211> 226
<212> PRT
<213> Streptomyces avermitilis

```
<400> 7
Met Thr Leu Gly Arg Pro Arg Arg Ser Ser Ala Asp Arg Pro Ala Pro
 1               5                  10                  15
Pro Ala Gly Ala Arg Ala Thr Ala Ala Gly Val Thr Val Arg Arg Leu
            20                  25                  30
Val Val Glu Gly Ala Val Glu Phe Thr Pro Thr Val Phe Pro Asp Glu
        35                  40                  45
Arg Gly Leu Phe Val Thr Pro Tyr Gln Glu Pro Val Leu Ser Glu Ala
    50                  55                  60
Val Gly His Arg Phe Pro Thr Ala Gln Thr Cys Gln Ser Val Ser Arg
65                  70                  75                  80
Arg Gly Val Val Arg Gly Val His Phe Thr Ala Thr Pro Pro Gly Gln
                85                  90                  95
Ala Lys Tyr Val His Cys Ala Arg Gly Arg Ala Leu Asp Phe Val Val
            100                 105                 110
Asp Leu Arg Thr Gly Ser Pro Thr Phe Gly Gln Trp Asp Ser Val Leu
        115                 120                 125
Leu Asp Gln Glu Arg Phe Arg Ser Val Tyr Leu Pro Ile Gly Val Gly
    130                 135                 140
His Ala Phe Val Ala Leu Glu Asp Asp Thr Ala Met Val Tyr Leu Met
145                 150                 155                 160
Ser Ser Gly Tyr Val Pro Gln Asn Glu His Ala Leu Ser Pro Glu Asp
                165                 170                 175
Pro Asp Leu Ala Leu Pro Leu Gly His His Leu Gly Arg Ala Pro Ile
            180                 185                 190
Leu Ser Glu Arg Gly Pro Ala Arg Ala Pro Thr Leu Gln Gln Ala Leu
        195                 200                 205
Arg Arg Gly Met Leu Pro Glu Tyr Arg Ala Ser Arg Ala Leu Asp Glu
    210                 215                 220
Lys Leu
225
```

<210> 8
<211> 464
<212> PRT
<213> Streptomyces avermitilis

```
<400> 8
Met Pro Thr Thr Pro Ser Pro Ala Pro Leu Thr Ala Arg His Asp Ala
 1               5                  10                  15
Ala Leu Pro Ala Cys Leu Ala Arg Ser Ala Ala Val Gly Asp Thr Gly
            20                  25                  30
Ala Arg Thr Ser Leu Asp Ala Phe Thr Gly Trp Trp Thr Arg Arg Ser
        35                  40                  45
Gly Ala His Arg Phe Arg Val Glu Arg Ile Pro Phe His Gly Met Asp
    50                  55                  60
```

```
Ala Trp Ser Phe His Pro Gly Thr Gly Asn Leu Ala His Arg Ser Gly
65                  70                  75                  80
Arg Phe Phe Ser Val Glu Gly Leu His Val Arg Gly Gly Glu Gln Pro
                85                  90                  95
Phe Pro Glu Trp Gln Gln Pro Ile Ile His Gln Pro Glu Ile Gly Ile
            100                 105                 110
Leu Gly Ile Leu Ala Lys Lys Phe Asp Gly Val Leu His Phe Leu Met
        115                 120                 125
Gln Ala Lys Met Glu Pro Gly Asn Ile Asn Leu Val Gln Leu Ser Pro
    130                 135                 140
Thr Val Gln Ala Thr Arg Ser Asn Tyr Thr Lys Val His Gly Gly Ala
145                 150                 155                 160
Ala Val Lys Tyr Leu Glu Tyr Phe Thr Gln Pro Arg Arg Ala Thr Val
                165                 170                 175
Val Val Asp Val Leu Gln Ser Glu His Gly Ala Trp Phe His Arg Lys
            180                 185                 190
Phe Asn Arg Asn Ile Val Val Glu Thr Asp Glu Asp Val Pro Leu Asp
        195                 200                 205
Asp Asp Phe Arg Trp Leu Thr Leu Gly Gln Ile Gly Glu Leu Met His
    210                 215                 220
Arg Asp Asn Leu Val Asn Met Asp Ala Arg Thr Val Leu Ala Cys Leu
225                 230                 235                 240
Pro Thr Pro Phe Asp Glu Pro Ala Ala Leu His Ser Asp Ala Glu Leu
                245                 250                 255
Leu Ser Trp Tyr Ala Ala Glu Arg Ser Arg His Ser Val His Ala Arg
            260                 265                 270
Arg Val Pro Leu Ala Gly Ile Pro Gly Trp Thr Thr Gly Ala Glu Ser
        275                 280                 285
Ile Ala His His Ala Asp Arg Tyr Phe Arg Val Val Ala Val Arg Val
    290                 295                 300
Glu Ala Ser Asn Arg Glu Val Ala Ala Trp Thr Gln Pro Leu Ile Glu
305                 310                 315                 320
Pro Cys Gly His Gly Ile Thr Ala Phe Leu Thr Arg Arg Ile Gly Gly
                325                 330                 335
Val Pro His Leu Leu Ala His Gly Arg Val Glu Gly Gly Phe Leu Asp
            340                 345                 350
Thr Ile Glu Leu Gly Pro Thr Val Gln Tyr Thr Pro Arg Asn Tyr Ala
        355                 360                 365
His Leu Thr Gly Pro Ala Arg Pro Arg Phe Leu Asp Leu Val Leu Glu
    370                 375                 380
Ala Ala Pro Asp Arg Ile Arg Tyr Ala Ala Val His Ser Glu Glu Gly
385                 390                 395                 400
Gly Arg Phe Leu His Ala Gln Ala Arg Tyr Leu Phe Val Glu Ala Asp
                405                 410                 415
Glu Ser Gln Ala Pro Asn Asp Pro Pro Pro Gly Tyr Arg Trp Cys Thr
            420                 425                 430
Pro Gly Gln Leu Thr Gln Leu Leu Arg Tyr Gly Arg Tyr Val Asn Val
        435                 440                 445
Gln Ala Arg Thr Leu Leu Ser Leu Leu Thr Thr Arg Ala Val Glu Leu
    450                 455                 460
```

<210> 9
<211> 257
<212> PRT
<213> Streptomyces avermitilis

<400> 9

```
Met Thr Glu Arg Glu Phe Thr Asp Pro Arg Ile Val Pro His Glu Ser
 1               5                  10                  15
Glu Gln Glu Arg Ala Ala Arg Glu Gln Leu Thr Lys Leu Leu Val Asp
            20                  25                  30
Ser Pro Ile Pro Pro Lys Tyr Leu Ile Asp Asn Leu Ser Val Tyr Met
```

```
        35                  40                  45
Arg Arg Asn Gln Leu Ala Asp Leu Leu Ser Met Asp Ala Leu Tyr Arg
    50                  55                  60
Met Leu Pro Glu Val Pro Gly Val Ile Met Glu Phe Gly Val Leu His
65                  70                  75                  80
Gly Arg His Leu Ala Thr Leu Thr Ala Leu Arg Ser Ile Tyr Glu Pro
                85                  90                  95
Tyr Asn Ser Leu Arg Arg Val Ile Gly Phe Asp Thr Phe Thr Gly Phe
            100                 105                 110
Pro Asp Ile Asp Glu Ala Asp Glu Val Ser Thr Ser Ala Val Pro Gly
        115                 120                 125
Arg Phe Ala Val Pro Asp Gly Glu Val Glu His Leu Arg Gln Val Leu
    130                 135                 140
Ala Ala His Glu Ala Asn Glu Pro Tyr Gly His Thr Gln Arg Ser Phe
145                 150                 155                 160
Val Val Gln Gly Asp Val Arg Glu Thr Val Pro Gln Tyr Leu Ala Glu
                165                 170                 175
His Pro His Thr Val Ile Ala Leu Ala Tyr Phe Asp Leu Asp Leu Tyr
            180                 185                 190
Arg Pro Thr Arg Glu Leu Leu Asp Val Ile Thr Pro His Leu Thr Arg
        195                 200                 205
Gly Ser Ile Leu Ala Phe Asp Glu Leu Thr His Pro Lys Trp Pro Gly
    210                 215                 220
Glu Thr Arg Ala Leu Ser Glu Ala Phe Gly Leu Asp His Ala Pro Leu
225                 230                 235                 240
Arg Gln Leu Pro Gly Arg Glu Pro Pro Val Ile Tyr Met Thr Trp Gly
                245                 250                 255
Ala
```

<210> 10
<211> 347
<212> PRT
<213> Streptomyces avermitilis

<400> 10

```
Met Met Pro Thr Thr Ala Glu Pro Ala Pro Val Gln Ser Asp Ser Asn
 1               5                  10                  15
Ser Ser Ala Pro Leu His Thr Glu Leu Gly Arg Thr Arg Leu Arg Ile
            20                  25                  30
Ser Arg Leu Ala Leu Gly Thr Val Asn Ile Gly Gly Arg Val Glu Glu
        35                  40                  45
Pro Glu Ala Arg Arg Leu Met Asp His Ala Leu Ala Gln Gly Ile Thr
    50                  55                  60
Leu Phe Asp Thr Ala Asn Thr Tyr Gly Trp Arg Val His Lys Gly Tyr
65                  70                  75                  80
Thr Glu Glu Val Ile Gly Arg Trp Leu Ala Asp Arg Pro Ala Arg Arg
                85                  90                  95
Glu Gln Val Val Leu Ala Thr Lys Val Gly Asp Pro Met Gly Ser Gly
            100                 105                 110
Pro Asn Asp His Gly Leu Ser Val Arg Asn Ile Val Ala Ala Cys Asp
        115                 120                 125
Ala Ser Leu Arg Arg Leu Arg Thr Asp Trp Ile Asp Leu Tyr Gln Leu
    130                 135                 140
His His Ile Asp Arg Arg Ala Gly Trp Asp Glu Val Trp Gln Ala Met
145                 150                 155                 160
Asp Leu Leu Ile Thr Gln Gly Lys Val Arg Tyr Val Gly Ser Ser Asn
                165                 170                 175
Phe Ala Gly Trp Asp Ile Ala Ser Ala Gln Glu Ala Ala Arg Arg Arg
            180                 185                 190
Asn Ala Leu Gly Leu Ala Ser Glu Gln Cys Val Tyr Asn Leu Val Thr
        195                 200                 205
```

```
Arg His Ala Glu Leu Glu Val Ile Pro Ala Ala Ser Ala Tyr Gly Val
    210                 215                 220
Gly Val Leu Val Trp Ser Pro Leu His Gly Gly Leu Leu Gly Gly Val
225                 230                 235                 240
Leu Arg Lys Thr Arg Glu Asn Thr Ala Val Lys Ser Ala Gln Gly Arg
                245                 250                 255
Ala Val Glu Ala Leu Glu His His Arg Thr Thr Ile Ala Ala Tyr Glu
            260                 265                 270
Asp Val Cys Ala Asp His Gly Leu Asp Pro Ala His Val Gly Met Ala
        275                 280                 285
Trp Val Leu Ser Arg Pro Gly Val Thr Gly Leu Val Ile Gly Pro Arg
    290                 295                 300
Thr Glu Gln His Val Asp Gly Ala Leu His Ala Leu Arg Thr Pro Leu
305                 310                 315                 320
Pro Glu Pro Val Leu Ala Arg Leu Glu Glu Leu Phe Pro Pro Val Gly
                325                 330                 335
Arg Gly Gly Ser Ala Pro Asp Ala Trp Leu Ser
            340                 345
```

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/20331

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/02, 21/04; C12N 9/00, 15/00, 5/00, 1/20

US CL : 536/23.1,23.2; 435/320.1, 325, 252.3,183

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1,23.2; 435/320.1, 325, 183

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST, MEDLINE, CAPLUS, EMBASE, BIOSIS
search terms: Douglas Macneil, James Occi, Keith Gewain, oleandrose, avermectin

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P --- Y,P	IKEDA, H. et al. Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in Steptomyces avermitilis. Proc. Natl. Acad. Sci. August 1999, Vol. 96, pages 9509-9514, entire document	1-3,5-7,11 --------- 8-10
X	WO 97/08323 A1 (CIBA-CEIGY AG) 06 March 1997, pages 37-41, see entire document.	1-3

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 06 SEPTEMBER 2000	Date of mailing of the international search report 05 OCT 2000
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Dorothea Lawrence for PETER TUNG Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)*